

# ViSNet: a scalable and accurate geometric deep learning potential for molecular dynamics simulation

Yusong Wang<sup>1,2†</sup>, Shaoning Li<sup>2†</sup>, Xinheng He<sup>3,4,2</sup>, Mingyu Li<sup>5,2</sup>, Zun Wang<sup>2</sup>, Nanning Zheng<sup>1</sup>, Bin Shao<sup>2\*</sup>, Tong Wang<sup>2\*</sup> and Tie-Yan Liu<sup>2</sup>

<sup>1</sup>Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, 710049, China.

<sup>2</sup>Microsoft Research, Beijing, 100080, China.

<sup>3</sup>The CAS Key Laboratory of Receptor Research and State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, 201203, China.

<sup>4</sup>University of Chinese Academy of Sciences, Beijing, 100049, China.

<sup>5</sup>Medicinal Chemistry and Bioinformatics Center, School of Medicine, Shanghai Jiaotong University, Shanghai, 200025, China.

\*Corresponding author(s). E-mail(s): [binshao@microsoft.com](mailto:binshao@microsoft.com) (B. S.);

[watong@microsoft.com](mailto:watong@microsoft.com) (T. W., Lead Contact);

†These authors contributed equally to this work.

## Abstract

Geometric deep learning has been revolutionizing the molecular dynamics simulation field over a decade. Although the state-of-the-art neural network models are approaching *ab initio* accuracy for energy and force prediction, insufficient utilization of geometric information and high computational costs hinder their applications in molecular dynamics simulations. Here we propose a deep learning potential, called ViSNet that sufficiently exploits directional information with low computational costs. ViSNet outperforms the state-of-the-art approaches on the molecules in the MD17 and revised MD17 datasets and achieves the best prediction scores for 11 of 12 quantum properties on QM9. Furthermore, ViSNet can scale to protein molecules containing hundreds of atoms and reach to *ab initio* accuracy without molecular segmentation. Through a series of evaluations and case studies, ViSNet can efficiently explore the conformational space and provide reasonable interpretability to map geometric representations to molecular structures.

**Keywords:** Geometric Deep Learning Potential; Equivariant Graph Neural Network; Molecular Dynamics Simulations

## 1 Introduction

Driven by the atomic forces calculated by potentials, molecular dynamics (MD) simulation

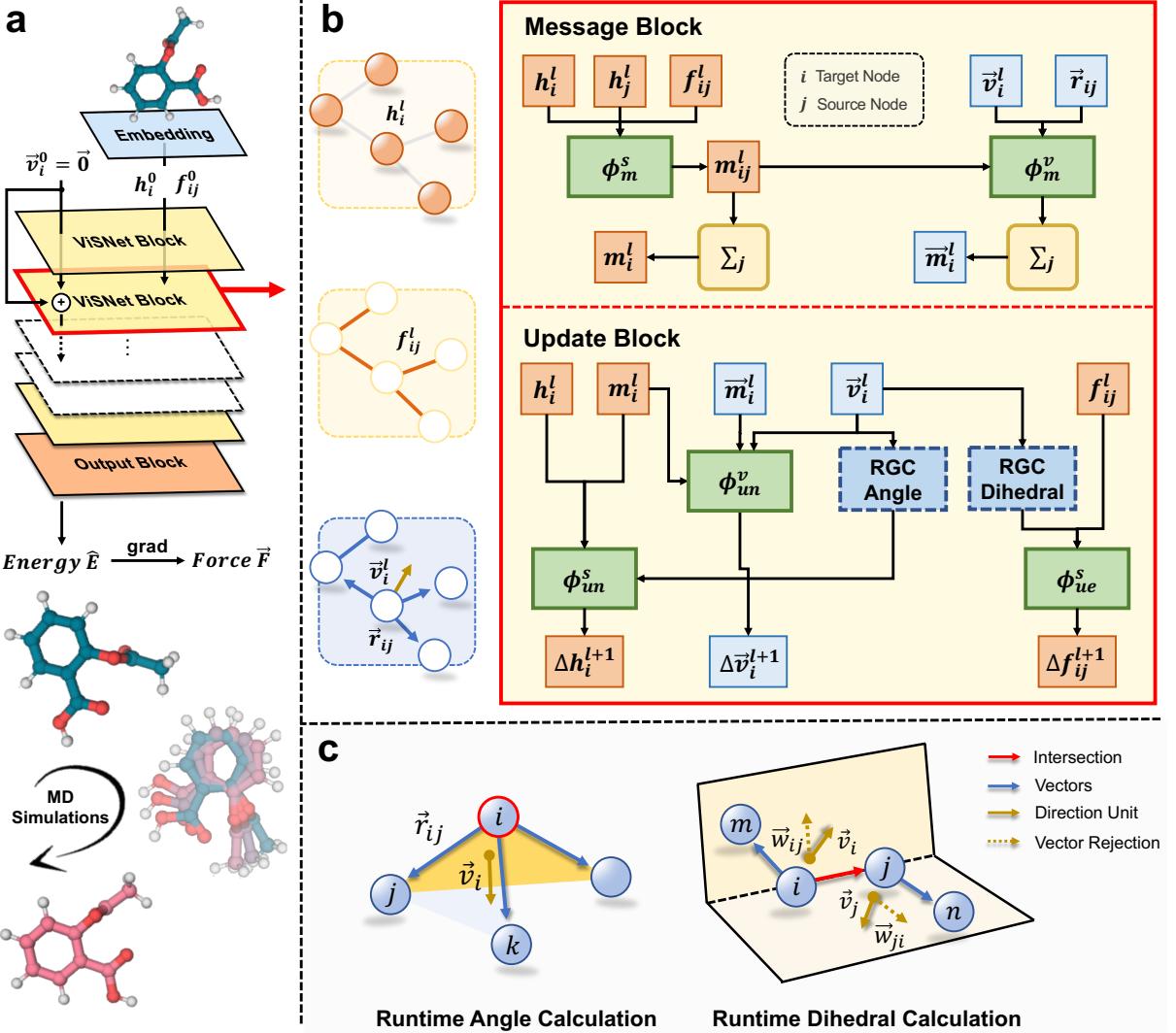
describes the kinetic and thermodynamic properties of a molecular system. It is widely used in physical, chemical, biological and pharmaceutical

fields [1–4]. *Ab initio* molecular dynamics simulations such as those driven by density functional theory (DFT) can accurately calculate energy and forces with a computational prohibitive cost, limiting its application to large molecular systems and long simulations [5, 6]. By contrast, classical MD employing empirical force fields [7] can do fast and suitable simulations as they’ve engendered for large systems while the quantum effect caused by electron movement cannot be captured and the force fields’ parameters are generally not transferable [8]. Due to powerful representation capabilities of machine learning (ML), ML potentials offer an alternative solution by learning from the reference data with *ab initio* accuracy and high computational efficiency [9, 10]. Behler and Parrinello first adopted descriptors to characterize the atomic local environment combining with a shallow multi-layer perception to learn the potential energy of molecules. In recent years, deep learning (DL) has demonstrated its powerful ability to learn from raw data without any hand-crafted features in many fields and thus DL potentials have attracted more and more attention. However, the inherent drawback of deep learning, which requires large amounts of data, has become a bottleneck for its application to more scenarios [12]. To alleviate the dependency on data for DL potentials, recent works have incorporated the inductive bias of symmetry into neural network design, termed as the geometric deep learning (GDL). Symmetry describes the conservation of physical laws, i.e., the unchanged physical properties with any transformations such as translations or rotations. Therefore, GDL can be extended to limited data scenarios without any data augmentation.

Equivariant graph neural network (EGNN) is one of the representative approaches in GDL, which has powerful capability to model molecular geometry [12–21]. A popular kind of EGNN potentials conduct equivariance from the group representation theory [12–15]. For example, the most recent study, NequIP [12] achieves the state-of-the-art performance on several molecular dynamics simulation datasets by leveraging high-order geometric tensors. Although these algorithms have solid mathematical guarantees and make full use of geometric information by high-order geometric tensors, operations in them such as Clebsch-Gordan product (CG-product) usually lead to

ultra-large computational overheads at an intolerable computational scale, which severely prohibits them from being applied for large molecules in practice [22]. To alleviate this limitation, another mainstream approach extracts geometric information explicitly to reduce the computational overheads and accelerate model training process. SchNet first introduces continuous-filter convolutions with rotational invariance [23]. DimeNet [16] and DimeNet++ [17] incorporate the angular information to improve the ability to model directional information. PaiNN [18] and Equivariant Transformer [19] further adopt vector embedding and scalarize the angular representation via inner product of the vector embedding itself, which extend SchNet and DimeNet from invariance to equivariance. Gasteiger et al., [20] employs dihedral information into DimeNet and proposes GemNet to promote directional modeling for molecules. On the one hand, the superior performance of these algorithms have proved the high effectiveness of encoding angle and dihedral information explicitly, which further leads to lower computational overheads by avoiding CG-product operations. On the other hand, operations designed in the previous studies for scalar and vector embeddings in the message passing module lack of sufficient utilization of such geometric information. This may hinder the model to well capture molecular representations and thus result in it not robust to various conformations during molecular dynamics simulation. Furthermore, it is worth noticing that as molecules become large, the number of angles and dihedrals increases dramatically, and thus the time consumption of explicit geometric extraction will also become neglectable.

In this study, we propose ViSNet (short for “Vector-Scalar interactive graph neural Network”), a scalable and accurate graph deep learning potential for molecular dynamics that significantly alleviate the dilemma between computational costs and sufficient utilization of geometric information. In ViSNet, we first proposed a Runtime Geometric Computation (RGC) strategy to extract and encode angular and dihedral information with linear computational complexity, significantly accelerating model training and inference as well as reducing the memory consumption. We then designed a novel and effective Vector-Scalar interactive equivariant Message Passing mechanism, termed as ViS-MP, to show how to make full



**Fig. 1 The architecture of ViSNet.** (a) The sketch of ViSNet. ViSNet embeds the 3D structures of molecules and extracts the geometric information through a series of ViSNet blocks and outputs the energy and forces through an output block. The forces predicted by ViSNet drive molecular dynamics simulation. (b) The flowchart of one ViSNet Block. A ViSNet block consists of two sub-blocks: i). Message Block; ii). Update Block. The two modules form an effective vector-scalar interactive message passing (ViS-MP) mechanism as illustrated by Eq. 5 to Eq. 9. Concretely, in the message block, the scalar messages  $m_{ij}$  and vector messages  $\vec{m}_{ij}$  are first obtained through message function  $\phi_m$  from node embedding  $h_i$ , edge embedding  $f_{ij}$ , relative position  $\vec{r}_{ij}$ , and direction unit  $\vec{v}_i$ . Then, they are aggregated to the target node  $i$ . In the update block,  $h_i$  is updated by the aggregated scalar message  $m_i$  and the output of RGC-Angle from  $\vec{v}_i$  through an update function  $\phi_{un}^s$ . Then,  $f_{ij}$  is updated by the output of RGC-Dihedral from  $\vec{v}_i$  and  $\vec{v}_j$  through an update function  $\phi_{ue}^s$ . The Runtime Geometry Calculation (RGC) strategy is explained in panel (c). Finally, the vector embedding  $\vec{v}_i$  is updated by both scalar and vector messages through an update function  $\phi_v^s$ . The overall design of ViS-MP aims to improve the interaction between scalar and vector embeddings. (c) An illustration on Runtime Geometry Calculation (RGC) strategy that extracts and calculates the directional geometric information with linear complexity. RGC comprises of two functions - Runtime Angle Calculation (RGC Angle) and Runtime Dihedral Calculation (RGC Dihedral). In RGC, the equivariant vector representation (termed as “direction unit”) for each node is designed to preserve its geometric information. Through rejections and inner products between two direction units (Eq. 1 - Eq. 4), the angle and dihedral information can be directly obtained with lower computational complexity, i.e.,  $\mathcal{O}(N)$ , for both angle and dihedral calculation.

use of geometric information by interacting vectorial hidden representations with scalar hidden representations. By incorporating these two modules, ViSNet combines the respective advantage of both types of EGNN, i.e., the high computational efficiency and sufficient utilization of geometric information. When comprehensively evaluated on some benchmarks, ViSNet outperforms all state-of-the-art algorithms on all molecules in MD17 and revised MD17 datasets and shows superior performance on QM9 dataset, indicating the powerful capability of molecular geometric representation. We then performed *ab initio* molecular dynamics simulations (AIMD) for each molecule on MD17 driven by ViSNet trained only with 0.7% of the data. The highly consistent interatomic distance distributions and the explored potential energy surfaces between AIMD and quantum simulation illustrate that ViSNet is genuinely data-efficient and can perform simulations with high fidelity. To further explore the scalability of ViSNet to large molecules, we built a full-atom MD dataset for the simplest protein Chignolin at DFT level that consists of 9543 different conformations of the 166-atom protein derived from replica exchange molecular dynamics and calculated by DFT. To the best of our knowledge, this is the first MD dataset for real-world full-atom proteins at DFT level. When evaluated on this dataset, ViSNet also achieved the best performance compared with DL potentials and empirical force fields. In addition, ViSNet exhibits reasonable interpretability to map geometric representation in ViSNet to molecular structures. The contributions of ViSNet can be summarized as follows:

- Proposing RGC module and ViS-MP mechanism to extract and exploit geometric information sufficiently with less computational costs and reasonable model interpretability.
- Achieving the state-of-the-art performance in various benchmarks for predicting energy, force and other quantum properties.
- Performing *ab initio* molecular dynamics simulations with high fidelity driven by ViSNet trained on limited data.
- Designing a full-atom protein MD dataset at DFT level and demonstrating ViSNet scaled from small organic molecules to proteins.

## 2 Results

### 2.1 Overview of ViSNet

ViSNet is a versatile GDL potential which predicts potential energy, atomic forces as well as various quantum chemical properties by taking atomic coordinates and numbers as inputs. As shown in Fig.1(a), the model is composed of an embedding block and multiple stacked ViSNet blocks, followed by an output block. The atomic number and coordinates are fed into the embedding block followed by ViSNet blocks to extract and encode geometric representations. The geometric representations are then used to predict molecular properties through the output block. It is worth noting that ViSNet is an energy-conserving potential, i.e., the predicted atomic forces are derived from the negative gradients of the potential energy with respect to the coordinates [11].

As shown in Fig.1(b), each ViSNet block consists of a message block and an update block. Both blocks are important parts of a novel vector-scalar interactive message passing mechanism, termed as ViS-MP. The rich geometric information embedded in messages is extracted by a runtime geometric calculation (RGC) strategy with linear complexity (Fig.1(c)). RGC strategy and ViS-MP mechanism are two key components of ViSNet and will be explained in detail in the following paragraphs. In addition, the glossary of notations used in ViSNet is shown in Supplementary Table 1.

**RGC: Runtime Geometry Calculation** The success of classical force fields shows that geometric features such as interatomic distances, angles, and dihedrals are essential to determine the total potential energy of molecules. The explicit extraction of invariant geometric representations in previous studies often suffer from a large amount of time or memory consumption during model training and inference. Given an atom, the calculation of angular information scales  $\mathcal{O}(N^2)$  with the number of neighboring atoms, while the computational complexity is even  $\mathcal{O}(N^3)$  for dihedrals [20]. To alleviate this problem, inspired by [18], we propose runtime geometry calculation (RGC), which uses an equivariant vector representation (termed as “direction unit”) for each node to preserve its geometric information. RGC directly calculates the geometric information from the direction unit which only sums the vectors from

the target node to its neighbors once. Therefore, the computational complexity can be reduced to  $\mathcal{O}(\mathcal{N})$ .

Considering the sub-structure of a toy molecule with four atoms shown in the Fig. 1(c), the angular information of the target node  $i$  could be conducted from the vector  $\vec{r}_{ij}$  as follows:

$$\vec{u}_{ij} = \frac{\vec{r}_{ij}}{\|\vec{r}_{ij}\|}, \quad \vec{v}_i = \sum_{j=1}^{N_i} \vec{u}_{ij} \quad (1)$$

$$\|\vec{v}_i\|^2 = \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} \langle \vec{u}_{ij}, \vec{u}_{ik} \rangle = \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} \cos \theta_{jik} \quad (2)$$

where  $\vec{r}_{ij}$  is the vector from node  $i$  to its neighboring node  $j$ ,  $\vec{u}_{ij}$  is the unit vector of  $\vec{r}_{ij}$ . Here, we propose the “direction unit”  $\vec{v}_i$  of node  $i$  as the sum of all unit vectors from node  $i$  to its all neighboring nodes  $j$ , where node  $i$  is the intersection of all unit vectors. As shown in Eq. 2, we calculate the inner product of direction unit  $\vec{v}_i$  of node  $i$  which represents the sum of inner products of unit vectors from node  $i$  to all its neighboring nodes. Combining with Eq. 1, the inner product of direction direction  $\vec{v}_i$  finally stands for the sum of cosine values of all angles formed by node  $i$  and any two of its neighboring nodes.

Similar to runtime angle calculation, we also calculate the vector rejection of the direction unit  $\vec{v}_i$  of node  $i$  and  $\vec{v}_j$  of node  $j$  on the vector  $\vec{u}_{ij}$  and  $\vec{u}_{ji}$ , respectively.

$$\begin{aligned} \vec{w}_{ij} &= \text{Rej}_{\vec{u}_{ij}}(\vec{v}_i) = \vec{v}_i - \langle \vec{v}_i, \vec{u}_{ij} \rangle \cdot \vec{u}_{ij} \\ &= \sum_{m=1}^{N_i} \text{Rej}_{\vec{u}_{ij}}(\vec{u}_{im}) \\ \vec{w}_{ji} &= \text{Rej}_{\vec{u}_{ji}}(\vec{v}_j) = \vec{v}_j - \langle \vec{v}_j, \vec{u}_{ji} \rangle \cdot \vec{u}_{ji} \\ &= \sum_{n=1}^{N_j} \text{Rej}_{\vec{u}_{ji}}(\vec{u}_{jn}) \end{aligned} \quad (3)$$

where  $\text{Rej}_{\vec{b}}(\vec{a})$  represents the vector component of  $\vec{a}$  perpendicular to  $\vec{b}$ , termed as the vector rejection.  $\vec{u}_{ij}$  and  $\vec{v}_i$  are defined in Eq. 1.  $\vec{w}_{ij}$  represents the sum of the vector rejection  $\text{Rej}_{\vec{u}_{ij}}(\vec{u}_{im})$  and  $\vec{w}_{ji}$  represents the sum of the vector rejection  $\text{Rej}_{\vec{u}_{ji}}(\vec{u}_{jn})$ . The inner product between  $\vec{w}_{ij}$  and  $\vec{w}_{ji}$  is then calculated to conduct dihedral

information of the axis of  $\vec{u}_{ij}$  as follows:

$$\begin{aligned} \langle \vec{w}_{ij}, \vec{w}_{ji} \rangle &= \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \left\langle \text{Rej}_{\vec{u}_{ij}}(\vec{u}_{im}), \text{Rej}_{\vec{u}_{ji}}(\vec{u}_{jn}) \right\rangle \\ &= \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \cos \varphi_{mijn} \end{aligned} \quad (4)$$

By calculating the inner product of the vector  $\vec{w}_{ij}$  with the vector  $\vec{w}_{ji}$ , we can obtain the sum of cosine values of all dihedrals  $\varphi_{mijn}$  with  $e_{ij}$  as the common rotation axis as shown in Fig. 1(c). Note that the directional unit is not restricted to Cartesian coordinates but can be extended to higher order tensors by spherical harmonics. We provide a proof about the rotational invariance of RGC strategy in the Supplementary.

**ViS-MP: Vector-Scalar interactive Message Passing** In order to make full use of geometric information and to enhance the interaction between scalars and vectors, we designed a universal vector-scalar interactive message passing mechanism with respect to the intersecting nodes and edges for angles and dihedrals, respectively. The key operations in ViS-MP are given as follows:

$$m_i^l = \sum_{j \in \mathcal{N}(i)} \phi_m^s(h_i^l, h_j^l, f_{ij}^l) \quad (5)$$

$$\vec{m}_i^l = \sum_{j \in \mathcal{N}(i)} \phi_m^v(m_{ij}^l, \vec{r}_{ij}, \vec{v}_j^l) \quad (6)$$

$$h_i^{l+1} = \phi_{un}^s(h_i^l, m_i^l, \langle \vec{v}_i^l, \vec{v}_i^l \rangle) \quad (7)$$

$$f_{ij}^{l+1} = \phi_{ue}^s(f_{ij}^l, \langle \text{Rej}_{\vec{r}_{ij}}(\vec{v}_i^l), \text{Rej}_{\vec{r}_{ji}}(\vec{v}_j^l) \rangle) \quad (8)$$

$$\vec{v}_i^{l+1} = \phi_{un}^v(\vec{v}_i^l, m_i^l, \vec{m}_i^l) \quad (9)$$

where  $h_i$  denotes the scalar embedding of node  $i$ ,  $f_{ij}$  stands for the edge feature between node  $i$  and node  $j$ .  $\vec{v}_i$  represents the embedding of direction unit mentioned in RGC. The superscript of variables indicate the index of the block that the variables belong to. ViS-MP extends the conventional message passing, aggregation, and update processes with vector-scalar interactions. Eq. 5 and Eq. 6 depict our message passing and aggregation processes. To be concrete, scalar messages  $m_{ij}$  incorporating scalar embedding  $h_j$ ,  $h_i$ , and  $f_{ij}$  are passed and then aggregated to node  $i$  through an message function  $\phi_m^s$  (Eq. 5). Similar operations are applied for vector messages  $\vec{m}_i^l$  of node

$i$  that incorporates scalar message  $m_{ij}$ , vector  $\vec{r}_{ij}$  and vector embedding  $\vec{v}_j$  (Eq. 6). Eq. 7 and Eq. 8 demonstrate the update processes.  $h_i$  is updated by the aggregated scalar message output  $m_i$  while the inner product of  $\vec{v}_i$  is updated through an update function  $\phi_{un}^s$ . Then  $\vec{f}_{ij}$  is updated by the inner product of the rejection of the vector embedding  $\vec{v}_i$  and  $\vec{v}_j$  through an update function  $\phi_{ue}^s$ . Finally the vector embedding  $\vec{v}_i$  is updated by both scalar and vector messages through an update function  $\phi_{un}^v$ . Notably, the non-linear functions for vectors, i.e.,  $\phi^v$  require to be equivariant. The detailed message and update functions can be found in the Methods section and Extended Data Fig. ??.

In summary, the geometric features are extracted by the inner products with the RGC strategy and the scalar and vector embeddings are cyclically updating each other in ViS-MP so as to learn a comprehensive geometric representation from molecular structures.

## 2.2 Accurate quantum chemical property predictions

We evaluated ViSNet on several prevailing benchmark datasets including MD17 [11, 24, 25], revised MD17 (termed as “rMD17”) [26], and QM9 [27] for energy, force, and other molecular property prediction. MD17 consists of the MD trajectories of 7 small organic molecules; the number of conformations in each molecule dataset ranges from 133,700 to 993,237. The dataset rMD17 is a reproduced version of MD17 with higher accuracy. QM9 consists of 12 kinds of quantum chemical properties of 133,385 small organic molecules with up to 9 heavy atoms. We compared ViSNet with the state-of-the-art algorithms, including the kernel-based algorithms FCHL19 [9] and GAP [10], the directional information-based algorithms SchNet [23], ANI [28], PhysNet [29], EGNN [30], ACE [31], DimeNet/DimeNet++ [16, 17], GemNet [20], PaiNN [18], and ET [19] and the group representation theory-based algorithms UNiTE [32] and NequIP [12]. The training details of ViSNet on each benchmark are described in the Methods section.

As shown in Table 1 and 2, it is remarkable that ViSNet outperformed the compared algorithms for all molecules with lowest energy and forces mean absolute errors (MAEs). Although

with only 950 samples of each kind of molecule for training and another 50 samples in the validation set, ViSNet still outperforms the kernel-based algorithms with a large margin, which indicates the equivariant model design in ViSNet captures geometric information efficiently and thus significantly alleviates the requirements of a large number of training samples. On the one hand, compared with PaiNN and ET, ViSNet incorporates more directional geometric information through RGC strategy, which contributes to performance gains. On the other hand, with similar angle and dihedral information adopted in GemNet, the superior performance of ViSNet indicates ViS-MP can better leverage geometric information during message passing. Furthermore, as shown in Extended Data Table 1, ViSNet also achieved the superior performance for quantum chemical property predictions on QM9. It outperformed the compared algorithms for 11 of 12 chemical properties and achieved the comparable result on the remaining property. Elaborated evaluations on various benchmarks confirmed the high prediction accuracy of ViSNet.

## 2.3 Efficient molecular dynamics simulations on MD17

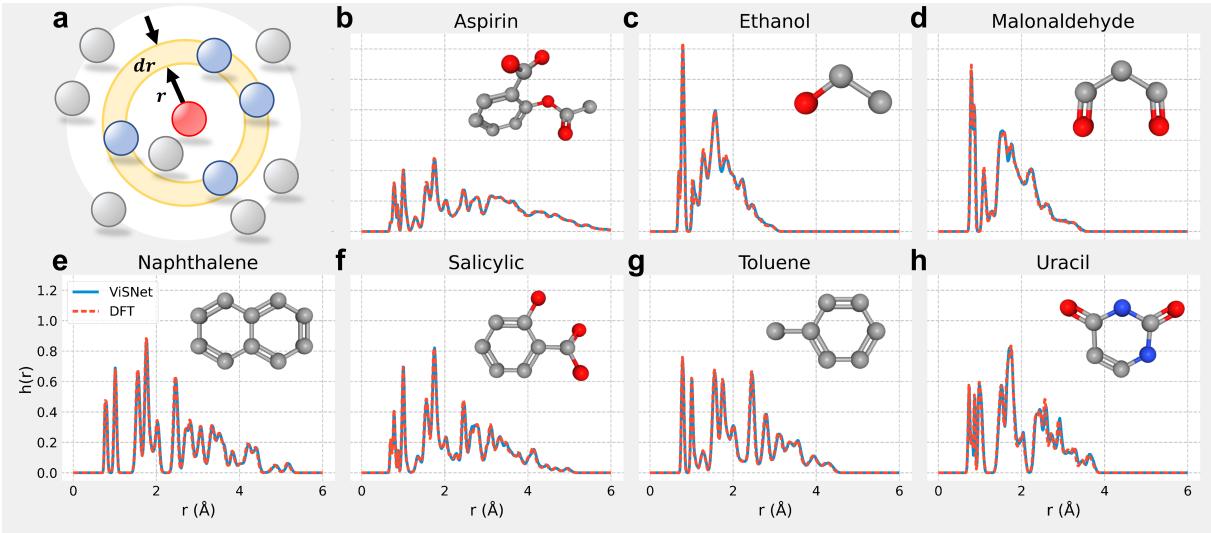
Most of the recently proposed methods are quite accurate in predicting potential energy and atomic forces for the conformations in a given test set. Molecular dynamics simulation is one of the important applications of the predicted potential energy and atomic forces. To evaluate ViSNet as the potential for *ab initio* molecular dynamics simulation, we incorporated ViSNet that trained only with 0.7% samples (i.e., 950 samples for model training) on MD17 into the ASE simulation framework [33] to perform *ab initio* MD simulations for all 7 kinds of organic molecules. All simulations are run with a time step  $\tau = 0.5$  fs under Berendsen thermostat with the other settings the same as those of the MD17 dataset. As shown in Fig. 2, we analyzed the interatomic distance distributions derived from both AIMD simulations with ViSNet as the potential and *ab initio* molecular dynamics simulations at DFT level for all 7 molecules, respectively. As shown in Fig. 2(a), the interatomic distance distribution  $h(r)$  is defined as the ensemble average of

**Table 1** Mean absolute errors (MAE) of energy (kcal/mol) and force (kcal/mol/Å) for 7 small organic molecules on MD17 compared with state-of-the-art algorithms. The best one in each category is highlighted in bold.

Molecule		SchNet	PhysNet	DimeNet	PaiNN	SpookyNet	ET	GemNet <sup>1</sup>	NequIP <sup>2</sup>	ViSNet
Aspirin	energy	0.37	0.230	0.204	0.167	0.151	0.123	-	0.131	<b>0.116</b>
	forces	1.35	0.605	0.499	0.338	0.258	0.253	0.217	0.184	<b>0.155</b>
Ethanol	energy	0.08	0.059	0.064	0.064	0.052	0.052	-	<b>0.051</b>	<b>0.051</b>
	forces	0.39	0.160	0.230	0.224	0.094	0.109	0.085	0.071	<b>0.060</b>
Malondialdehyde	energy	0.13	0.094	0.104	0.091	0.079	0.077	-	0.076	<b>0.075</b>
	forces	0.66	0.319	0.383	0.319	0.167	0.169	0.155	0.129	<b>0.100</b>
Naphthalene	energy	0.16	0.142	0.122	0.116	0.116	<b>0.085</b>	-	0.113	<b>0.085</b>
	forces	0.58	0.310	0.215	0.077	0.089	0.061	0.051	<b>0.039</b>	<b>0.039</b>
Salicylic Acid	energy	0.20	0.126	0.134	0.116	0.114	0.093	-	0.106	<b>0.092</b>
	forces	0.85	0.337	0.374	0.195	0.180	0.129	0.125	0.090	<b>0.084</b>
Toluene	energy	0.12	0.100	0.102	0.095	0.094	<b>0.074</b>	-	0.092	<b>0.074</b>
	forces	0.57	0.191	0.216	0.094	0.087	0.067	0.060	0.046	<b>0.039</b>
Uracil	energy	0.14	0.109	0.115	0.106	0.105	<b>0.095</b>	-	0.104	<b>0.095</b>
	forces	0.56	0.218	0.301	0.139	0.119	0.095	0.097	0.076	<b>0.062</b>

<sup>1</sup> The best results are reported among four variants of GemNet.

<sup>2</sup> NequIP only shows the results with  $l = 3$ .



**Fig. 2** The interatomic distance distributions of MD simulations driven by ViSNet and DFT. (a) An illustration about the atomic density at a radius  $r$  with the arbitrary atom as the center. The interatomic distance distribution is defined as the ensemble average of atomic density. (b) to (h) The interatomic distance distributions comparison between simulations by ViSNet and DFT for all seven organic molecules in MD17. The curve of ViSNet is shown using a solid blue line, while the dashed orange line is used for DFT curve. The structures of the corresponding molecules are shown in the upper right corner.

atomic density at a radius  $r$  [25]. Fig. 2(b-h) illustrate the distributions derived from ViSNet are very close to those generated by DFT. We also compared the potential energy surfaces sampled

by ViSNet and DFT for these molecules, respectively (Extended Data Fig. ??). The consistent potential energy surfaces suggest that ViSNet can

**Table 2** Mean absolute errors (MAE) of energy (kcal/mol) and force (kcal/mol/Å) for 10 small organic molecules on rMD17 compared with state-of-the-art algorithms. The best one in each category is highlighted in bold.

Molecule		FCHL19	UNiTE <sup>3</sup>	GAP	ANI	ACE	GemNet <sup>1</sup>	NequIP <sup>1</sup>	ViSNet <sup>2</sup>
Aspirin	energy	0.143	0.055	0.408	0.383	0.141	-	0.0530	<b>0.0445</b>
	forces	0.482	0.175	1.035	0.936	0.413	0.2191	0.1891	<b>0.1520</b>
Azobenzene	energy	0.065	0.025	0.196	0.367	0.083	-	0.0161	<b>0.0156</b>
	forces	0.249	0.097	0.565	0.816	0.251	-	0.0669	<b>0.0585</b>
Benzene	energy	0.007	0.002	0.017	0.076	0.0009	-	0.0009	<b>0.0007</b>
	forces	0.060	0.017	0.138	0.231	0.012	0.0115	0.0069	<b>0.0056</b>
Ethanol	energy	0.021	0.014	0.081	0.058	0.028	-	0.0092	<b>0.0078</b>
	forces	0.143	0.085	0.417	0.309	0.168	0.083	0.0646	<b>0.0522</b>
Malonaldehyde	energy	0.035	0.025	0.111	0.106	0.039	-	0.0184	<b>0.0132</b>
	forces	0.235	0.152	0.609	0.565	0.256	0.1522	0.1176	<b>0.0893</b>
Naphthalene	energy	0.028	0.011	0.088	0.261	0.021	-	<b>0.0046</b>	0.0057
	forces	0.150	0.060	0.380	0.673	0.118	0.0438	0.0300	<b>0.0291</b>
Paracetamol	energy	0.067	0.044	0.196	0.265	0.092	-	0.0323	<b>0.0258</b>
	forces	0.281	0.164	0.666	0.701	0.293	-	0.1361	<b>0.1029</b>
Salicylic acid	energy	0.042	0.017	0.129	0.212	0.042	-	0.0161	<b>0.0161</b>
	forces	0.219	0.088	0.570	0.685	0.214	0.1222	0.0922	<b>0.0795</b>
Toluene	energy	0.037	0.010	0.092	0.178	0.025	-	0.0069	<b>0.0059</b>
	forces	0.203	0.058	0.410	0.560	0.150	0.0507	0.0369	<b>0.0264</b>
Uracil	energy	0.014	0.013	0.069	0.118	0.025	-	0.0092	<b>0.0069</b>
	forces	0.097	0.088	0.406	0.493	0.152	0.0876	0.0669	<b>0.0495</b>

<sup>1</sup> The best results are reported among four variants of GemNet and four orders  $l \in \{0, 1, 2, 3\}$  of NequIP.

<sup>2</sup> ViSNet can achieve better results with longer convergence time.

<sup>3</sup> For a fair comparison, the “direct learning” results without any extra input are compared.

well recover the kinetic properties and the conformational space from the simulation trajectories, indicating the usefulness of ViSNet for real molecular dynamics simulation. Furthermore, compared with the prohibitive computational cost of DFT, ViSNet dramatically saves the computational time by 2-3 orders of magnitude (Extended Data Fig. ?? and Supplementary Table 2). These results demonstrate that with only a few of training samples, ViSNet can act as the potential to perform high-fidelity molecular dynamics simulations with much less computational cost.

## 2.4 Scaling to real-world full-atom proteins

We further applied ViSNet to real-world proteins to explore its scalability from small organic molecules to large biomolecules. Considering that the time complexity of DFT roughly scales  $\mathcal{O}(N^3)$  with the number of atoms, we employed the simplest protein *Chignolin* with 166 atoms to build an MD dataset at DFT level for model training and evaluation. For data generation, we ran a 80 ns Replica Exchange Molecular Dynamics (REMD) simulation [34, 35] to sample various folding and unfolding states of *Chignolin*. As a result, 9,543 representative conformations were collected and the energy and forces on nuclei were calculated

by Gaussian 16 software package [36]. To the best of our knowledge, this is the first MD dataset for real-world full-atom proteins at the DFT level. The data generation process is elaborated in the Methods section. We split the *Chignolin* dataset as training, validation, and test sets by the ratio of 8:1:1. We trained ViSNet as well as the models with the best performance in the evaluations elaborated in Section 2.2 including ET, NequIP, and GemNet on the *Chignolin* dataset with their default settings on Tesla V100 GPUs. During model training, GemNet failed due to out of GPU memory even though the batch size is set to 1 while NequIP suffered from under-fitting with its default hyperparameters on *Chignolin* dataset. ViSNet and ET could successfully be trained and compared with molecular mechanics (MM). The DFT results were used as the ground truth. Fig. 3(a) shows the free energy landscape of *Chignolin* sampled by REMD and depicted by  $d_{Y2-G6}$  (the distance between mainchain O on Y2 and main-chain N on G6) and  $d_{E4-T7}$  (the distance between mainchain O on E4 and mainchain N on T7). The concentrated energy basin on the left shows the folded state and the scattered energy basin on the right shows unfolded state. We picked six representative structures in the low potential energy regions with both folded and unfolded states and selected some intermediate states with high potential energy colored cyan or blue. We visualized the energy predictions for the six representative structures, and ViSNet produced a significantly better estimation of the potential energy than both ET trained on the same dataset and MM with empirical force fields did. Fig. 3(b) to (d) show the correlations between the predicted energies by ViSNet, ET, MM, and the ground truth values given by DFT for the conformations in the test set. ViSNet achieved the lowest MAE and the highest  $R^2$  score. Similar results can be seen in the force correlations shown in the Supplementary Fig. 1. These results suggest that ViSNet has the ability to scale to the real-world proteins with a small training set and achieves superior accuracy and efficiency.

## 2.5 Ablation study

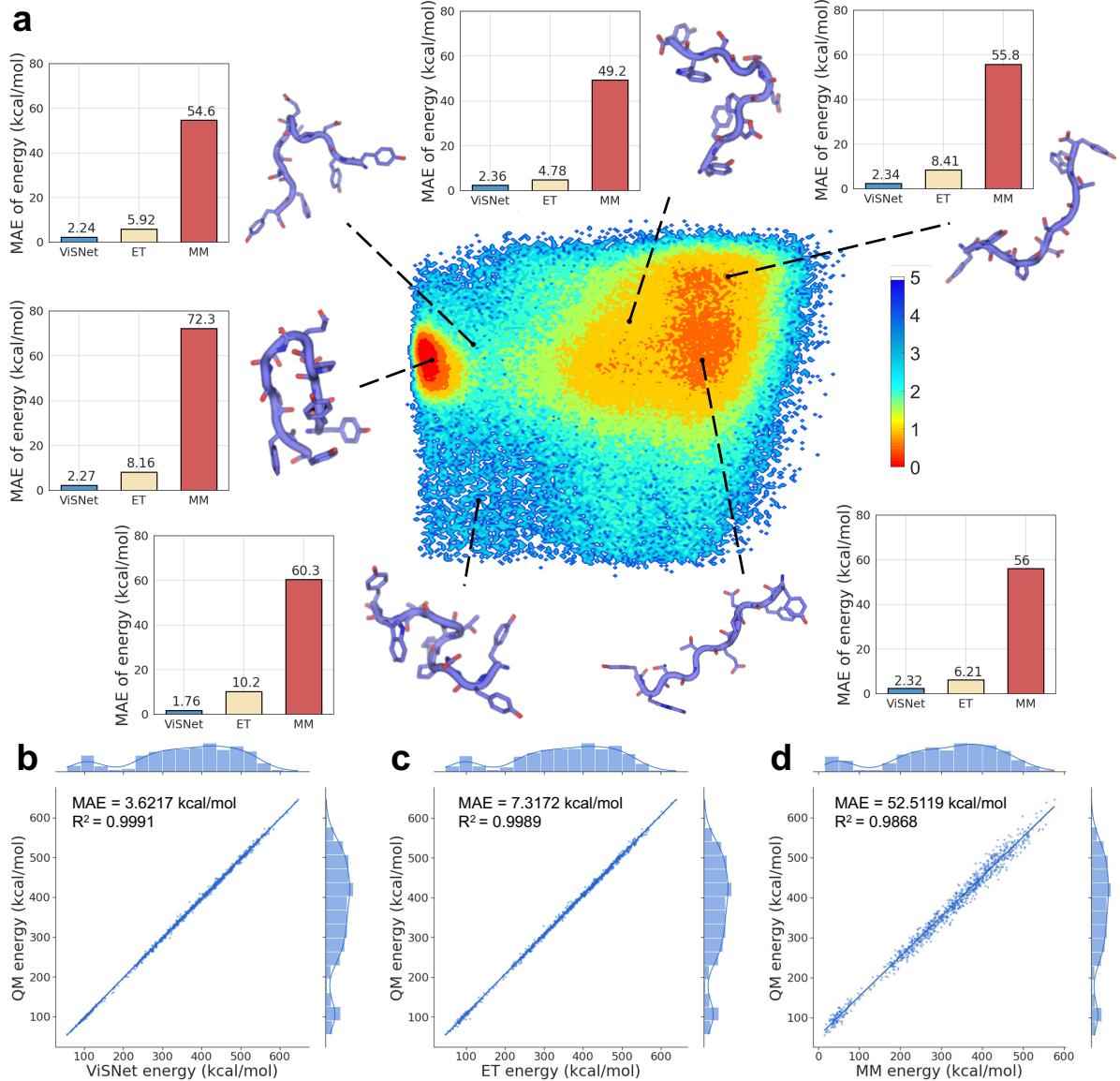
To further explore where the performance gains of ViSNet come from, we conducted a comprehensive

ablation study. Specifically, we excluded the runtime angle calculation (w/o A), runtime dihedral calculation (w/o D), and both of them (w/o A&D) in ViSNet, in order to evaluate the usefulness of each part. We designed some model variants with different message passing mechanisms based on ViS-MP for scalar and vector interaction. ViSNet-N directly aggregates the dihedral information to intersecting nodes, and ViSNet-T leverages another form of dihedral calculation. The details of these model variants are elaborated in Supplementary. The results of the ablation study are shown in Extended Data Table 2. Based on the results, we can see that both kinds of directional geometric information are useful and the dihedral information contributes a little bit more to the final performance. Furthermore, the significant performance drop from ViSNet-N and ViSNet-T further validate the effectiveness of ViS-MP mechanism.

## 2.6 Interpretability of ViSNet on molecular structures

Prior works have shown the effectiveness of incorporating geometric features, such as angles. However, they seldom make an explanation for how it improves the expressiveness of GNNs. To this end, we illustrate a reasonable model interpretability of ViSNet by mapping the angle representations derived from inner product of direction units in the model to the atoms in the molecular structure. We aim to bridge the gap between geometric representation in ViSNet and molecular structures. We visualized the embeddings after the inner product of direction units  $\langle \vec{v}_i, \vec{v}_i \rangle$  extracted from 50 aspirin samples on the validation set. The high-dimensional embeddings were reduced to 2-dimensional space using T-SNE [37] and then clustered using DBSCAN [38] without the prior of the number of clusters.

Fig. 4 exhibits the clustering results of nodes' embeddings after the inner product of their corresponding direction units. We further map the clustered nodes to the atoms of aspirin chemical structure. Interestingly, the embeddings for these nodes could be distinctly gathered into several clusters shown in different colors. For example, although carbon atom  $C_{11}$  and carbon atom  $C_{12}$  possess different positions and connect with different atoms, their inner product  $\langle \vec{v}_i, \vec{v}_i \rangle$  are clustered

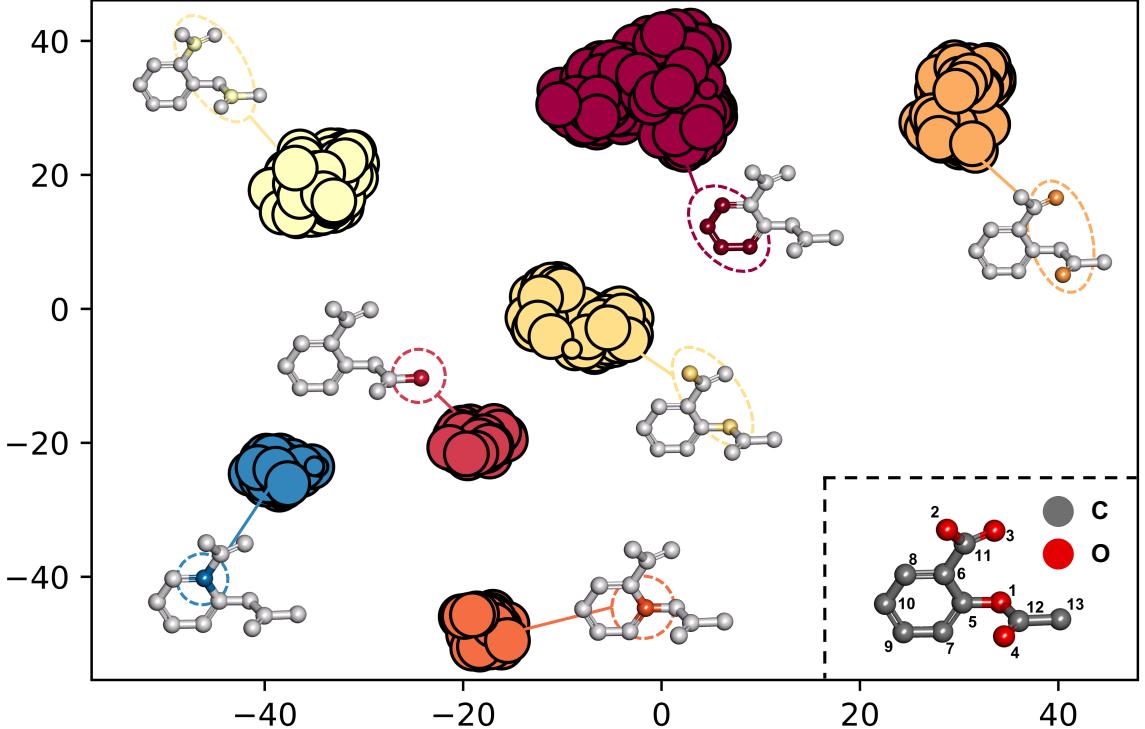


**Fig. 3 Visualization of the energy landscape of *Chignolin* and evaluations of energy prediction by ViSNet, ET, and molecular mechanics. (a)** The energy landscape of *Chignolin* sampled by REMD. The x-axis of the landscape is the distance between mainchain O on Y2 and mainchain N on G6, while the y-axis is the distance between mainchain O on E4 and mainchain N on T7. 6 representative structures were then selected for visualization. Each structure is shown as cartoon and residues are depicted in sticks. The histograms show the mean absolute error (MAE) between the energy difference predicted/calculated by ViSNet, ET, MM, and the ground truth calculated by DFT on the corresponding structure. **(b) to (d)** The energy correlations on the test dataset between the ground truth calculated by DFT and the predictions made by ViSNet, ET, and molecular mechanics. The corresponding distributions of energy predictions or calculations as well as the ground truth are shown in each panel.

into the same class for holding similar substructures ( $\{C_{11} - O_2O_3C_6\}$  and  $\{C_{12} - O_1O_4C_{13}\}$ ). To summarize, ViSNet can discriminate different molecular substructures in the embedding space.

### 3 Discussion and conclusion

We propose ViSNet, a novel geometric deep learning potential for molecular dynamics simulation. The group representation theory based methods and the directional information based methods are



**Fig. 4 Visualization and model interpretability of ViSNet.** Clusters of nodes' embeddings after the inner product of the direction units  $\langle \vec{v}_i, \vec{v}_i \rangle$ . The  $\langle \vec{v}_i, \vec{v}_i \rangle$  represents angle representations with the intersecting node  $i$  as the vertex. The atoms in the chemical structure of aspirin corresponding to each cluster are colored with the same color of the cluster, while the remaining atoms are colored light gray. A chemical structure of Aspirin and the indices of atoms are illustrated in the bottom right region. Carbon and oxygen atoms are colored dark grey and red, respectively. The hydrogen atoms are omitted in both the clustering results and the chemical structure of aspirin for simplification.

two mainstream classes of geometric deep learning potentials to enforce SE(3) equivariance [20]. ViSNet takes the advantages from both sides in designing RGC strategy and ViS-MP mechanism. On the one hand, the RGC strategy explicitly extracts and exploits the directional geometric information with computationally lightweight operations, making the model training and inference fast. On the other hand, ViS-MP employs a series of effective and efficient vector-scalar interactive operations, leading to the full use of the geometric information. Furthermore, according to the many-body expansion theory [39–41], the potential energy of the whole system equals to the potential of each single atom plus the energy corrections from two-bodies to many-bodies. Most of the previous studies model the truncated energy correction terms hierarchically with  $k$ -hop information via stacking  $k$  message passing blocks. Different from these approaches, ViSNet encodes the triplet and quadruplet interactions in a single

block, which empowers the model to have much more powerful representation ability. In addition, considering that angle and dihedral are important potential terms in empirical force fields, the interpretability of the operations in the RGC strategy provides some insights in constructing hybrid force fields by combining empirical terms with deep learning.

Besides predicting energy, force, and chemical properties with high accuracy, performing molecular dynamics simulations with *ab initio* accuracy at the cost of empirical force field is a grand challenge. ViSNet proves its usefulness in real-world *ab initio* molecular dynamics simulations with less computational costs and the ability of scaling to large molecules such as proteins. Extending ViSNet to support larger and more complex molecular systems will be our future research direction.

## 4 Methods

### 4.1 Detailed operations and modules in ViSNet

ViSNet predicts the molecular properties (e.g., energy  $\hat{E}$ , forces  $\vec{F} \in \mathbb{R}^{N \times 3}$ , dipole moment  $\mu$ ) from the current states of atoms, including the atomic positions  $X \in \mathbb{R}^{N \times 3}$  and atomic numbers  $Z \in \mathbb{N}^N$ . The architecture of the proposed ViSNet is shown in Fig. 1. The overall design of ViSNet follows the vector-scalar interactive message passing as illustrated from Eq. 5 - Eq. 8. First, an embedding block encodes the atom numbers and edge distances into the embedding space. Then, a series of ViSNet blocks update the node-wise scalar and vector representations based on their interactions. A residual connection is placed between two ViSNet blocks. Finally, stacked corresponding gated equivariant blocks proposed by [18] are attached to the output block for specific molecular property prediction.

**The Embedding block** ViSNet expands the direct node and edge embedding with their neighbors. It first embeds atomic chemical symbol  $z_i$ , and calculates the edge representation whose distances within the cutoff through radial basis functions (RBF). Then the initial embedding of the atom  $i$ , its 1-hop neighbors  $j$  and the directly connected edge  $e_{ij}$  within cutoff are fused together as the initial node embedding  $h_i^0$  and edge embedding  $f_{ij}^0$ . In summary, the embedding block is given by:

$$h_i^0, f_{ij}^0 = \text{Embedding Block}(z_i, z_j, e_{ij}), \quad j \in \mathcal{N}(i) \quad (10)$$

$\mathcal{N}(i)$  denotes the set of 1-hop neighboring nodes of node  $i$ , and  $j$  is one of its neighbors. The embedding process is elaborated in Supplementary. The initial vector embedding  $\vec{v}_i$  is set to  $\vec{0}$ . The vector embeddings  $\vec{v}$  are projected into the embedding space by following [18];  $\vec{v} \in \mathbb{R}^{N \times 3 \times F}$  and  $F$  is the size of hidden dimension. The advantage of such projection is to assign a unique high-dimensional representation for each embedding to discriminate from each other. Further discussions on its effectiveness and interpretability are given in the Results section.

**The Scalar2Vec module** In the Scalar2Vec module, the vector embedding  $\vec{v}$  is updated by both the scalar messages derived from node and

edge scalar embeddings (Eq. 5) and the vector messages with inherent geometric information (Eq. 6). The message of each atom is calculated through an Edge-Fusion Graph Attention module, which fuses the node and edge embeddings and computes the attention scores. The fusion of the node and edge embeddings could be the concatenation operation, Hadamard product, or adding a learnable bias [44]. We leverage the Hadamard product and the *vanilla* multi-head attention mechanism borrowed from Transformer [45] for edge-node fusion.

Following [19], we pass the fused representations through a nonlinear activation function as shown in Eq. 11. The value ( $V$ ) in the attention mechanism is also fused by edge features before being multiplied by attention scores weighted by a cosine cutoff as shown in Eq. 12,

$$\alpha_{ij}^l = \sigma \left( (W_Q^l h_i^l) \left( W_K^l h_j^l \odot \text{Dense}_K^l(f_{ij}^l) \right)^T \right) \quad (11)$$

$$m_{ij}^l = \alpha_{ij}^l \cdot \phi(\|\vec{r}_{ij}\|) \cdot \left( W_V^l h_j^l \odot \text{Dense}_V^l(f_{ij}^l) \right) \quad (12)$$

where  $l \in \{0, 1, 2, \dots, L\}$  is the index of block,  $\sigma$  denotes the activation function (SiLU in this paper),  $W$  is the learnable weight matrix,  $\odot$  represents the Hadamard product,  $\phi(\cdot)$  denotes the cosine cutoff and  $\text{Dense}(\cdot)$  refers to one learnable weight matrix with activation function. For brevity, we omit the learnable bias for linear transformation on scalar embedding in equations, and there is no bias for vector embedding to ensure the universal equivariance.

Then, the computed  $m_{ij}^l$  is used to produce the geometric messages  $\vec{m}_{ij}^l$  for vectors:

$$\vec{m}_{ij}^l = \left( \text{Dense}_u^l(m_{ij}^l) \odot \vec{u}_{ij} \right) + \left( \text{Dense}_v^l(m_{ij}^l) \odot \vec{v}_j^l \right) \quad (13)$$

And the vector embedding  $\vec{v}^l$  is updated by:

$$m_i^l = \sum_{j \in \mathcal{N}(i)} m_{ij}^l, \quad \vec{m}_i^l = \sum_{j \in \mathcal{N}(i)} \vec{m}_{ij}^l \quad (14)$$

$$\Delta \vec{v}_i^{l+1} = \vec{m}_i^l + W_{\text{vm}}^l m_i^l \odot W_{\text{v}}^l \vec{v}_i^l \quad (15)$$

**The Vec2Scalar module** In the Vec2Scalar module, the node embedding  $h_i^l$  and edge embedding  $f_{ij}^l$  are updated by the geometric information extracted by the RGC strategy, i.e., angles (Eq.

7) and dihedrals (Eq. 8), respectively. The residual node embedding  $\Delta h_i^{l+1}$ , is calculated by a Hadamard product between the runtime angle information and the aggregated scalar messages with a gated residual connection:

$$\Delta h_i^{l+1} = \langle W_t^l \vec{v}_i^l, W_s^l \vec{v}_i^l \rangle \odot W_{\text{Angle}}^l m_i^l + W_{\text{res}}^l m_i^l \quad (16)$$

To compute the residual edge embedding  $\Delta f_{ij}^{l+1}$ , we perform the Hadamard product of the runtime dihedral information with the transformed edge embedding:

$$\begin{aligned} \Delta f_{ij}^{l+1} = & \left\langle \text{Rej}_{\vec{r}_{ij}}(W_{Rt}^l \vec{v}_i^l), \text{Rej}_{\vec{r}_{ji}}(W_{Rs}^l \vec{v}_j^l) \right\rangle \odot \\ & \text{Dense}_{\text{Dihedral}}^l(f_{ij}^l) \end{aligned} \quad (17)$$

After the residual hidden representations are calculated, we add them to the original input of block  $l$  and feed them to the next block.

**The output block** Following PaiNN [18], we update the scalar embedding and vector embedding of nodes with multiple gated equivariant blocks:

$$t_i^l = \text{Dense}_{o_2}^l(\|W_{o_1}^l \vec{v}_i^l\|, h_i^l) \quad (18)$$

$$h_i^{l+1} = W_{o_3}^l t_i^l \quad (19)$$

$$\vec{v}_i^{l+1} = W_{o_4}^l \vec{v}_i^l \odot W_{o_5}^l t_i^l \quad (20)$$

where  $[\cdot, \cdot]$  is the tensor concatenation operation. The final scalar embedding  $h_i^L \in \mathbb{R}^{N \times 1}$  and vector embedding  $\vec{v}_i^L \in \mathbb{R}^{N \times 3 \times 1}$  are used to predict various molecular properties.

On QM9, the molecular dipole is calculated as follows:

$$\mu = \left\| \sum_{i=1}^N \vec{v}_i^L + h_i^L (\vec{r}_i - \vec{r}_c) \right\| \quad (21)$$

where  $\vec{r}_c$  denotes the center of mass. Similarly, for the prediction of electronic spatial extent  $\langle R^2 \rangle$ , we use the following equation:

$$\langle R^2 \rangle = \sum_{i=1}^N h_i^L \|\vec{r}_i - \vec{r}_c\|^2 \quad (22)$$

For the remaining 10 properties  $y$ , we simply aggregate the final scalar embedding of nodes as

follows:

$$y = \sum_{i=1}^N h_i^L \quad (23)$$

For models trained on the molecular dynamics datasets including MD17, revised MD17, and *Chignolin*, the total potential energy is obtained as the sum of the final scalar embedding of the nodes. As an energy-conserving potential, the forces are then calculated using the negative gradients of the predicted total potential energy with respect to the atomic coordinates:

$$E = \sum_{i=1}^N h_i^L \quad (24)$$

$$\vec{F}_i = -\nabla_i E \quad (25)$$

## 4.2 The design of *Chignolin* dataset

The initial structure for replica exchange molecular dynamics (REMD) simulations was derived from protein data bank (PDB ID: 5AWL)[46]. Water molecules in the crystal structure were removed. Then, FF19SB force field [7] was applied to describe the atomic interactions for *Chignolin* in generalized Born implicit solvent model. A second modification of the Bondi Van der Waals radii set was used in the solvent model[47]. The program makeCHIR\_RST in Amber 20 was applied to create chiral restraint file during REMD simulation to maintain the chiral property at a high temperature. The system at the beginning encountered a minimization process of 500 steepest descent and 500 conjugate gradient cycles. After energy minimization, 200 ps of equilibration runs at 300 K, 400 K, 500 K, 600 K, 700 K, 800 K, 900 K, 1000 K were applied to the system with random initial velocities. The final structure of equilibration were used for REMD simulations at the corresponding temperatures. Each single replica in the production ran last 2 ps and then was exchanged to the neighbouring temperature. The exchange happened 5,000 times in each production run, and we had 8 replica temperatures, which led to a total simulation time of 80 ns. The sampling interval of each simulation trajectory was 0.4 ps so the trajectory had 200,000 points. We evenly picked 10,000 points from the REMD trajectory to generate the input file for Gaussian 16 [36]. The potential energy and the atomic forces

for each conformation were calculated with M06-2X functional and 6-31G\* basis. The integration grid was set to *superfine* precision.

Finally, 9,543 SCF converged conformations with the total potential energy and atomic forces were recruited in the *Chignolin* dataset. As shown in Supplementary Fig. 2, the distribution of the total energy ranges from -2,831,076.155 kcal/mol to -2,830,477.983 kcal/mol and some representative conformations are also picked for visualization. Note that the total energy does not show a normal distribution, but has two peaks corresponding to the folded and unfolded states of *Chignolin*, which increases the difficulty for model training on the dataset. The distributions of forces with respect to three axis ( $x, y, z$ ) as well as the distributions of the magnitude are shown in Supplementary Fig. 3. Other statistical information is available in Supplementary Table 3.

### 4.3 Dataset splitting schemes

For the QM9 dataset, we randomly split it into 110,000 samples as the train set, 10,000 samples as the validation set, and the rest as the test set by following the previous studies [18, 19].

To evaluate the effectiveness of ViSNet to simulation data, ViSNet was trained on MD17 and rMD17 with a limited data setting, which consists of only 950 uniformly sampled conformations for model training and 50 conformations for validation for each molecule.

Furthermore, the whole *Chignolin* dataset was randomly split into 80%, 10%, and 10% as the training, validation, and test datasets. Six representative conformations were picked from the test set for illustration.

### 4.4 Experimental settings

For the QM9 dataset, we adopted a batch size of 32 and a learning rate of 1e-4 for all the properties. The mean squared error (MSE) loss was used for model training. For the molecular dynamic dataset including MD17, rMD17, and *Chignolin*, we leveraged a combined MSE loss for energy and force prediction. The weight of energy loss was set to 0.05 for MD17 and rMD17, 0.2 for *Chignolin*. The weight of forces loss was set to 0.95 for MD17 and rMD17, 0.8 for Chignolin. The batch size was set to 4 and the learning rate was chosen from 2e-4, 3e-4, 4e-4 for different molecules. The cutoff was

set to 5 for small molecules in QM9, MD17, and rMD17 and changed to 4 for *Chignolin* in order to reduce the number of edges in the molecular graphs. We used the learning rate decay if the validation loss stopped decreasing. The patience was set to 15 epochs for QM9, and 30 epochs for MD17, rMD17, and *Chignolin*. The learning rate decay factor was set to 0.8 for these models. We also adopted the early stopping strategy to prevent over-fitting. The ViSNet model trained on the molecular dynamic datasets had 9 hidden layers and the embedding dimension was set to 256. We used a larger model for QM9 dataset, i.e., the embedding dimension changed to 512. More details about the hyperparameters of ViSNet can be found in Supplementary Table 4. Experiments were conducted on NVIDIA 32G-V100 GPUs.

## Author contributions

T. W. conceived and designed the study. S. L., Y. W. and T. W. carried out algorithm design. Y. W., S. L., X. H., M. L. and Z. W. carried out experiments, evaluations, analysis and visualization. Y. W. and S. L. wrote the original manuscript. T. W., X. H., Z. W. and B. S revised the manuscript. N. Z. and T. L. contributed to writing. All authors reviewed the final manuscript.

## References

- [1] Chow, E., Klepeis, J., Rendleman, C., Dror, R. & Shaw, D. 9.6 new technologies for molecular dynamics simulations. *Edward H. Egelman, editor. Comprehensive Biophysics. Amsterdam: Elsevier* 86–104 (2012).
- [2] Singh, S. & Singh, V. K. Molecular dynamics simulation: methods and application. In *Frontiers in protein structure, function, and dynamics*, 213–238 (Springer, 2020).
- [3] Lu, S. *et al.* Activation pathway of a g protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. *Nature Communications* **12**, 1–15 (2021).
- [4] Li, Y. *et al.* Exploring the regulatory function of the n-terminal domain of sars-cov-2 spike

- protein through molecular dynamics simulation. *Advanced theory and simulations* **4**, 2100152 (2021).
- [5] Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review* **140**, A1133 (1965).
- [6] Marx, D. & Hutter, J. *Ab initio molecular dynamics: basic theory and advanced methods* (Cambridge University Press, 2009).
- [7] Tian, C. *et al.* ff19sb: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *Journal of chemical theory and computation* **16**, 528–552 (2019).
- [8] Wang, H., Zhang, L., Han, J. & Weinan, E. Deepmd-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Computer Physics Communications* **228**, 178–184 (2018).
- [9] Christensen, A. S., Bratholm, L. A., Faber, F. A. & Anatole von Lilienfeld, O. Fchl revisited: Faster and more accurate quantum machine learning. *The Journal of chemical physics* **152**, 044107 (2020).
- [10] Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical review letters* **104**, 136403 (2010).
- [11] Chmiela, S., Sauceda, H. E., Müller, K.-R. & Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature communications* **9**, 1–10 (2018).
- [12] Batzner, S. *et al.* E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications* **13**, 1–11 (2022).
- [13] Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E. & Welling, M. Geometric and physical quantities improve e (3) equivariant message passing. *International Conference on Learning Representations* (2022).
- [14] Hutchinson, M. J. *et al.* Lietransformer: Equivariant self-attention for lie groups. In *International Conference on Machine Learning*, 4533–4543 (PMLR, 2021).
- [15] Fuchs, F., Worrall, D., Fischer, V. & Welling, M. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems* **33**, 1970–1981 (2020).
- [16] Gasteiger, J., Groß, J. & Günemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations* (2019).
- [17] Gasteiger, J., Giri, S., Margraf, J. T. & Günemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *Advances in Neural Information Processing Systems* (2020).
- [18] Schütt, K., Unke, O. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, 9377–9388 (PMLR, 2021).
- [19] Thölke, P. & De Fabritiis, G. Torchmd-net: Equivariant transformers for neural network based molecular potentials. *The International Conference on Learning Representations* (2022).
- [20] Gasteiger, J., Becker, F. & Günemann, S. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems* **34**, 6790–6802 (2021).
- [21] Unke, O. T. *et al.* Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature communications* **12**, 1–14 (2021).
- [22] Han, J., Rong, Y., Xu, T. & Huang, W. Geometrically equivariant graph neural networks: A survey. *arXiv preprint arXiv:2202.07230* (2022).

- [23] Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. Schnet—a deep learning architecture for molecules and materials. *The Journal of Chemical Physics* **148**, 241722 (2018).
- [24] Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nature communications* **8**, 1–8 (2017).
- [25] Chmiela, S. *et al.* Machine learning of accurate energy-conserving molecular force fields. *Science advances* **3**, e1603015 (2017).
- [26] Christensen, A. S. & Von Lilienfeld, O. A. On the role of gradients for machine learning of molecular energies and forces. *Machine Learning: Science and Technology* **1**, 045018 (2020).
- [27] Ramakrishnan, R., Dral, P. O., Rupp, M. & Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data* **1**, 1–7 (2014).
- [28] Smith, J. S., Isayev, O. & Roitberg, A. E. Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chemical science* **8**, 3192–3203 (2017).
- [29] Unke, O. T. & Meuwly, M. Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of chemical theory and computation* **15**, 3678–3693 (2019).
- [30] Satorras, V. G., Hoogeboom, E. & Welling, M. E (n) equivariant graph neural networks. In *International Conference on Machine Learning*, 9323–9332 (PMLR, 2021).
- [31] Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B* **99**, 014104 (2019).
- [32] Qiao, Z. *et al.* Informing geometric deep learning with electronic interactions to accelerate quantum chemistry. *Proceedings of the National Academy of Sciences* **119**, e2205221119 (2022).
- [33] Larsen, A. H. *et al.* The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter* **29**, 273002 (2017).
- [34] Qi, R., Wei, G., Ma, B. & Nussinov, R. Replica exchange molecular dynamics: A practical application protocol with solutions to common problems and a peptide aggregation and self-assembly example. In *Peptide self-assembly*, 101–119 (Springer, 2018).
- [35] Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters* **314**, 141–151 (1999).
- [36] Frisch, M. J. *et al.* Gaussian~16 Revision C.01 (2016). Gaussian Inc. Wallingford CT.
- [37] Van der Maaten, L. & Hinton, G. Visualizing data using t-sne. *Journal of machine learning research* **9** (2008).
- [38] Ester, M., Kriegel, H.-P., Sander, J., Xu, X. *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, vol. 96, 226–231 (1996).
- [39] Nesbet, R. Atomic Bethe-Goldstone equations. III. correlation energies of ground states of Be, B, C, N, O, F, and Ne. *Physical Review* **175**, 2 (1968).
- [40] Hankins, D., Moskowitz, J. & Stillinger, F. Water molecule interactions. *The Journal of Chemical Physics* **53**, 4544–4554 (1970).
- [41] Gordon, M. S., Fedorov, D. G., Pruitt, S. R. & Slipchenko, L. V. Fragmentation methods: A route to accurate calculations on large systems. *Chemical reviews* **112**, 632–672 (2012).
- [42] Behler, J. Representing potential energy surfaces by high-dimensional neural network potentials. *Journal of Physics: Condensed Matter* **26**, 183001 (2014).
- [43] Stocker, S., Gasteiger, J., Becker, F., Günemann, S. & Margraf, J. T. How robust are modern graph neural network potentials

- in long and hot molecular dynamics simulations? (2022).
- [44] Ying, C. *et al.* Do transformers really perform badly for graph representation? *Advances in Neural Information Processing Systems* **34** (2021).
  - [45] Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems* **30** (2017).
  - [46] Honda, S. *et al.* Crystal structure of a ten-amino acid protein. *Journal of the American Chemical Society* **130**, 15327–15331 (2008).
  - [47] Onufriev, A., Bashford, D. & Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **55**, 383–394 (2004).

**Extended Data Table 1** Mean absolute errors (MAE) of 12 kinds of molecular properties on QM9 compared with state-of-the-art algorithms. The best one in each category is highlighted in bold.

Target	Unit	SchNet	PhysNet	EGNN	DimeNet++	Cormorant	PaiNN	ET	ViSNet
$\mu$	$D$	0.033	0.0529	0.029	0.0297	0.038	0.012	0.011	<b>0.010</b>
$\alpha$	$a_0^m$	0.235	0.0615	0.071	0.0435	0.085	0.045	0.059	<b>0.0411</b>
$\epsilon_{HOMO}$	$meV$	41	32.9	29	24.6	34	27.6	20.3	<b>17.3</b>
$\epsilon_{LUMO}$	$meV$	34	24.7	25	19.5	38	20.4	17.5	<b>14.8</b>
$\Delta\epsilon$	$meV$	63	42.5	48	32.6	61	45.7	36.1	<b>31.7</b>
$\langle R^2 \rangle$	$a_0^2$	0.073	0.765	0.106	0.331	0.961	0.066	0.033	<b>0.030</b>
$ZPVE$	$meV$	1.7	1.39	1.55	<b>1.21</b>	2.027	1.28	1.84	1.56
$U_0$	$meV$	14	8.15	11	6.32	22	5.85	6.15	<b>4.23</b>
$U$	$meV$	19	8.34	12	6.28	21	5.83	6.38	<b>4.25</b>
$H$	$meV$	14	8.42	12	6.53	21	5.98	6.16	<b>4.52</b>
$G$	$meV$	14	9.4	12	7.56	20	7.35	7.62	<b>5.86</b>
$C_v$	$\frac{\text{cal}}{\text{mol K}}$	0.033	0.028	0.031	<b>0.023</b>	0.026	0.024	0.026	<b>0.023</b>

**Extended Data Table 2** Ablation study of ViSNet on aspirin in MD17 dataset. ViSNet is compared with its variants without runtime angle calculation (w/o A), without runtime dihedral calculation (w/o D), and neither of them (w/o A&D). ViSNet-N and ViSNet-T are two variants with different message passing mechanisms from ViS-MP (see Supplementary for more details). The best results are shown in bold.

	energy	forces
ViSNet	<b>0.116</b>	<b>0.155</b>
w/o A	0.121	0.174
w/o D	0.124	0.224
w/o A&D	0.289	0.654
VisNet-N	0.123	0.224
VisNet-T	0.136	0.281