Computers in Biology and Medicine 176 (2024) 108531

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Carbon-based molecular properties efficiently predicted by deep learning-based quantum chemical simulation with large language models

Haoyu Wang [a,b,1], Bin Chen [a,d,1,*], Hangling Sun [c], Yuxuan Zhang [a]

[a] University of Shanghai for Science and Technology, Shanghai, China
[b] School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China
[c] Hengtu Imalligent Technology (Shanghai) Co., Ltd., Shanghai, China
[d] School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Keywords:
Carbon-based energy
Molecular conformation
Quantum chemistry
Deep learning
Large language model

ABSTRACT

The prediction of thermodynamic properties of carbon-based molecules based on their geometrical conformation using fluctuation and density functional theories has achieved great success in the field of energy chemistry, while the excessive computational cost provides both opportunities and challenges for the integration of machine learning. In this work, a deep learning-based quantum chemical prediction model was constructed for efficient prediction of thermodynamic properties of carbon-based molecules. We constructed a novel framework — encoding the 3D information into a large language model (LLM), which in turn generates a 2D SMILES string, while embedding a learnable encoding designed to preserve the integrity of the original 3D information, providing better structural information for the model. Additionally, we have designed an equivariant learning module to encompass representations of conformations and feature learning for conformational sampling. This framework aims to predict thermodynamic properties more accurately than learning from 2D topology alone, while providing faster computational speeds than conventional simulations. By combining machine learning and quantum chemistry, we pioneer efficient practical applications in the field of energy chemistry. Our model advances the integration of data-driven and physics-based modeling to unlock novel insights into carbon-based molecules.

1. Introduction

Understanding the molecular properties of carbon-based fuels can deepen our insight into their thermochemical transformation mechanisms, allowing for optimized and controlled reaction processes [1]. In recent years, quantum chemical calculations have been extensively used to predict the properties of carbon-based fuels. These methods encode the complex three-dimensional structures of molecules [2], enabling precise calculations and predictions of molecular attributes such as energy, electronic structure, and spectral signals, demonstrating significant potential [3]. Wavefunction-based post hartree fock methods and density functional theory-based Kohn–Sham equations require substantial computational resources with larger basis sets, restricting their application to large-scale, high-precision numerical calculations. This often leads to simplifying models by focusing on representative small molecular fragments [4].

Vibrational analysis is essential for calculating thermodynamic properties (enthalpy, entropy, free energy), which involves accurately calculating the Hessian matrix. This process is time-consuming for medium-sized systems (30–80 atoms), especially when only first-order analytical derivatives are available, which are obtained through first-order finite difference methods [5]. Consequently, an increasing number of researchers are turning to machine learning (ML) approaches. ML offers the potential for building more efficient models for calculating and predicting molecular properties, with accuracy comparable to various numerical solutions based on the Schrödinger equation, but with significantly improved efficiency. Neural networks, in particular, are a highly flexible [6] and unbiased class of mathematical functions capable of approximating any real-valued function with arbitrary precision [7].

Recently, graph neural networks (GNNs) have shown promising paradigm shifts by modeling molecules as graphs and learning from 2D topological connectivity patterns [8]. Various GNN architectures have developed specifically for molecular graph inputs [9], achieving strong performance on predictions solely from 2D information [10]. However, these approaches still lack complete incorporation of 3D structural details which can be critical for capturing chemical complexity in Fig. 2.

* Corresponding author at: University of Shanghai for Science and Technology, Shanghai, China.
E-mail addresses: ambityuki@gmail.com (H. Wang), chenbin1933@163.com (B. Chen).
[1] This is to indicate the equal contribution.

https://doi.org/10.1016/j.compbiomed.2024.108531
Received 11 February 2024; Received in revised form 21 April 2024; Accepted 28 April 2024
Available online 1 May 2024

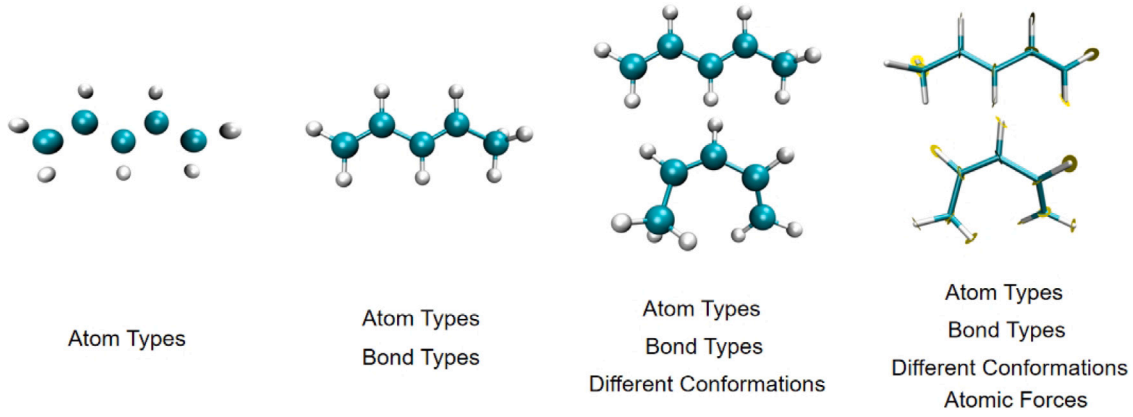

Fig. 1. Useful information from 3D molecular structure. The 3D structure of a molecule provides informative local geometric descriptors including bond lengths between atom pairs, bond angles formed by connected atom triplets, and dihedral angles defined by bonded atom quadruplets. Capturing these intricate details helps encode valuable conformational knowledge.

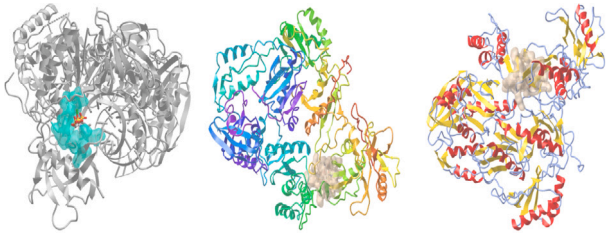

Fig. 2. The chemical simulation system exhibits intricate molecular structures with discernible correlations and distinct chemical bonds. Human interpretation relies on color labeling for differentiation, while machines can efficiently discern key features without this added step.

For example, small fluctuational changes in bond lengths and angles can drastically affect molecular properties through electronic structure alterations [11] in Fig. 1. While topology-based GNNs have made significant advances [12], explicitly encoding 3D geometric knowledge remains imperative for unlocking chemical prediction capabilities [13].

On the other hand, recent breakthroughs in natural language processing have given rise to a paradigm-shifting class of large pre-trained language models, such as generative pre-trained transformer (GPT) [14]. By leveraging self-supervised learning on massive textual corpora, these models have demonstrated unparalleled capacity for knowledge acquisition and few-shot generalization abilities across textual domains. Unlike images or graphs, sequential text representing molecules as SMILES strings can provide an information-rich data format amenable for language models like GPT [15]. Moreover, transforming 3D molecular graphs into sequential embeddings could enable leveraging both structural knowledge and language modeling capabilities [16]. However, effectively conveying chemical complexity to language models for property predictions remains filled with open challenges. For instance, critical 3D geometric details may be lost when converting conformations to sequences. Additionally, it remains unclear if language models can distinguish between distinct data types like equilibrium conformers versus non-equilibrium dynamical conformations.

Recent studies have shown that equivariant networks have emerged as effective architectures for learning symmetric patterns in data like molecules [17], but their differentiation between distinct data types like static conformers versus dynamical conformations is still lacking [18]. To address this question, we design a novel equivariant learning module to encode differences between equilibrium conformers and off-equilibrium conformations sampled from dynamics. Specifically, the module leveragesSteerable CNNs to extract equivariant features adaptive to rotations and permutations. We then investigate using transformer architecture to capture long-range dependencies based on the equivariant representations. By visualizing the attention weights, we analyze whether the model focuses on different atoms when seeing static conformers versus dynamical conformations. The designed architecture not only achieves improved performance on property predictions, but also provides interpretability on the model's understanding of molecular data.

In this work, we propose a novelty framework called 3D molecular structure enhanced training (**3D MolSE**) to combine the strengths of graph neural networks and language models for molecular property prediction. Specifically, we first leverage 3D graph networks to obtain informative molecular representations by pre-training on 3D molecular graphs. The learned structural embeddings are then transformed into sequences and fed into pre-trained transformers to predict target properties after fine-tuning. To retain and enhance 3D geometric information, we design tailored positional encodings for the graph embeddings. We use large language models to generate 2D SMILES strings as additional inputs to the transformer. Moreover, we have designed an equivariant learning module with the aim of considering and encompassing the feature learning regarding the learned representation of conformers versus conformations sampled from molecular dynamics or normal modes. The main contributions are summarized as follows.

- We propose a novelty framework that combines 3D graph neural network and LLM. The pre-trained module is to obtain molecular structural representations, which are then transformed into sequences and fed into LLM for downstream prediction.
- We propose an approach that combines improved encodings for input sequences, both fixed and learnable, with the utilization of LLM to generate 2D SMILES strings of molecules. This approach enhances the capture of structural information and provides additional molecular structure information.
- We design a specialized equivariant learning module to differentiate between conformers and conformations. The transformer architecture facilitates the analysis of attention weights, enhancing interpretability in the model.
- Through extensive experiments on diverse public benchmarks, our model effectively extracts 3D structural knowledge, surpassing the limitations of learning solely from 2D topology.

The remainder of the paper is organized as follows: Section 2 provides an overview of related work, while Section 3 presents the datasets and the proposed framework. In Section 4, we present the experimental findings, and in Section 5, we discuss the results of the investigations. Finally, we conclude our work in Section 6.

2. Related work

Predicting molecular properties from 3D structural data while bypassing expensive computations remains an active area of research. Significant progress has been made in two relevant directions — graph neural networks modeling 2D topology and generative language models leveraging big data. Equivariant networks for molecular modeling have also been effectively applied, enhancing predictive capacities in molecular property prediction.

2.1. Graph neural networks for molecules

In recent years, graph neural networks (GNNs) have emerged as a powerful paradigm for learning informative molecular representations [19]. Message passing neural networks (MPNNs) [20] in particular have achieved remarkable success by operating on atoms and bonds to model quantum interactions [21]. By propagating neighborhood information along molecular topology, MPNNs have shown state-of-the-art performance on diverse downstream tasks. However, significant challenges still remain [22,23] in advancing architectures to fully capture 3D conformational intricacies [24]. While recent variants have incorporated additional inputs like interatomic distances [25], the explicit encoding of precise geometric details including torsional angles and chirality still lacks [26]. Attempts to utilize spherical convolutions for local 3D property predictions show early promise but face restrictions in transferability across chemical environments.

The key opportunity lies in progressing pure topology-based networks to intrinsically learn from 3D structures. Purposeful injection of conformational biases and coordinate modeling capacities into neural networks operators could significantly boost their quantum mechanistic awareness [27]. Recent trends also point to hybrid physics-infused architectures [28] that unify message passing with equivariant transformations adapted to molecular symmetries. Overall, effectively bridging neural networks frameworks with elaborate 3D induction mechanisms remains imperative for transcending current limitations [29] and advancing molecular understanding to address pressing chemical demands.

2.2. Generative language models

Recent breakthroughs in natural language processing, especially large generative models pre-trained on massive textual corpora, could unlock new paradigms for chemical machine learning. Models like GPT [30] have demonstrated remarkable few-shot generalization capacities across diverse NLP domains. Preliminary studies have shown the feasibility of fine-tuning such models on molecular SMILES strings for basic property predictions [31]. However, systematically harnessing the multi-task knowledge transfer capabilities of large language models for accurate and scalable chemical modeling remains largely unexplored. Kenneth et al.[32] developing specialized tokenization and encoding strategies for handling domain data types like graphs and geometries, Hakime et al.[33] architectural adaptations encapsulating equivariances and invariances intrinsic to molecular energies and forces also warrant investigation. Furthermore, Ratul et al.[34] comprehensively encoding molecular structure-function relationships spanning 3D shapes, 2D connectivity, and quantum interactions poses additional modeling demands.

Daniel et al.[35] strategically bridging graph networks and language models via purpose-built intermediate representations. Encoding conformations as invariant or equivariant sequences [36] could serve as informative data flow conduits. Multi-levelinductive biases reflecting geometric, topological and electronic symmetries can further aid domain transfer [37]. Exploring these avenues by introducing tailored model integrations and encoding innovations could unlock new horizons for scalable, accurate and interpretable chemical predictions.

2.3. Equivariant networks for molecular modeling

Equivariant neural networks have recently emerged as effective learning frameworks by incorporating transformation symmetries within model architectures. By ensuring layers transform equivariantly under rotations and permutations, these networks can capture intrinsic patterns in molecular data. Works like 3D steerable networks [38] and SE(3)-Transformers [39] have shown improved performance by being equivariant to rigid motions and clashes. However, applications to molecular modeling remain scarce [40] except for preliminary studies on small datasets [41].

Challenges still exist in scaling equivariant networks for chemical graphs and effectively fusing them with sequential modeling [42]. Stephan et al. [43] analyze if and how these specialized architectures can distinguish between distinct molecular data types like static conformers versus dynamical conformations [44]. Questions around interpretability [45] also persist around how much equivariant networks uniquely focus on atomic neighborhoods when predicting overall molecular behaviors [46]. Incorporating known physical constraints and geometric relationships within model architectures aligned with chemical domains shows great yet untapped potential.

3. Materials and methods

3.1. Materials and chemicals

The QM9 dataset [47] contains DFT-optimized molecular structures and electronic properties calculated at the B3LYP/6-31G(2df, *p*) level of theory for 134k small organic molecules made up of CHONF atoms in Table 1. The equilibrium geometries correspond to local energy minima on the potential energy surface. Each molecule has only one low-energy conformer structure.

The dataset provides Cartesian coordinates specifying the 3D arrangement of atoms along with various ground truth quantum properties determined from solving the electronic Schrödinger equation. Target values include the highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbital energies, excitation energies, atomization energies, heat capacities, spatial extents of electron density, and more. QM9 encompasses a diverse chemical space of pharmaceutical-like organic compounds up to 9 heavy atoms. QM9 is split into 80% training, 10% validation and 10% testing sets for benchmarking. This static single-conformer dataset allows analysis of model performance on quantum predictions from genuine DFT-optimized 3D molecular structures.

To supplement the model, PW1 datasets were automatically extracted using cheminformatics workflows. These validated PW1 datasets provide spatial mappings of the charge distribution around molecules, which strongly dictate intermolecular interaction and reactivity. The automated extraction methodology ensures consistent large-scale dataset construction without manual bottleneck. In total, we have produced high-quality PW1 data for over 50,000 compounds across various structure families. This data-driven approach will facilitate robust molecular property modeling across a diverse chemical space.

$$E_{gap} = E_{LUMO} - E_{HOMO} \tag{1}$$

where E_{gap} is the HOMO-LUMO gap, E_{LUMO} is the energy of the lowest unoccupied molecular orbital, and E_{HOMO} is the energy of the highest occupied molecular orbital. As per Koopmans' theorem, E_{LUMO} correlates with the vertical electron affinity while E_{HOMO} correlates with the vertical ionization potential. Therefore, the HOMO-LUMO gap indicates chemical reactivity. Furthermore, chemical softness is inversely proportional to E_{gap}.

$$S \propto \frac{1}{E_{gap}} \tag{2}$$

Table 1
Quantum chemical properties on QM9 dataset.

Task	Property description	Unit
ϵ_{HOMO}	Highest occupied molecular orbital	eV
ϵ_{LUMO}	Lowest unoccupied molecular orbital	eV
μ	Molecular dipole moment	D
α	Isotropic electronic polarizability	a_0^3
U_0	Atomization energy at 0 K	eV
U	Atomization energy at 298.15 K	eV
$\Delta\epsilon$	Gap between HOMO and LUMO	eV
$\langle R^2 \rangle$	Average electronic spatial extent	a_0^3
ZPE	Zero-point vibrational energy	eV
H	Atomization enthalpy at 298.15 K	eV
G	Atomization free energy at 298.15 K	eV
c_v	Heat capacity at 298.15 K	cal/mol K

So a smaller HOMO-LUMO gap generally corresponds to higher chemical softness and reactivity. The HOMO-LUMO gap is thus a useful quantum chemical descriptor for estimating molecular reactivity trends. Some reference data in PW1 Dataset is provided for consideration at the link.[2]

3.2. Experimental setup and devices

The experiments were performed using PyTorch deep learning framework and RDKit cheminformatics toolkit. Training was conducted on RTX 4090 GPU with 24 GB RAM. Key hyperparameters are summarized in Table 2.

The 3D molecular graph network consists of 4 interaction blocks with hidden dimension 128, interleaved with BatchNorm and dropout regularization at a rate of 0.1. The output node and graph representations are projected to dimension 64 before sequence transformation. The transformer encoder contains 4 layers with 16 attention heads. The feedforward sublayers have dimension 512 with GeLU activations. Dropout of 0.1 is applied on the attention and output layers.

During 3D pre-training, each batch comprises 256 molecular graphs sampled randomly. The model is trained for 300 epochs with early stopping based on validation loss. Up to 80 conformers per molecule are generated on-the-fly for learning geometric tasks. For QM9 fine-tuning, batches contain 256 molecules constructed from raw XYZ coordinates [11]. The downstream model is optimized for 500 epochs using cosine decay learning rate scheduler. Predictions are made on the fixed test set and evaluated by mean absolute error between predicted and DFT-calculated target property values. To assess model performance on predicting quantum chemical properties of molecules, some key regression evaluation metrics are utilized.

$$F1\text{-}Mean = \frac{1}{n}\sum_{i=1}^{n}\frac{2PR_i}{P+R_i} \tag{3}$$

where P and R_i are precision and recall for sample i. Harmonic mean of precision and recall across samples. Higher indicates better performance.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

where true positives (TP) is the number of positive samples correctly predicted as positive. False negatives (FN) is the number of positive samples incorrectly predicted as negative. True negatives (TN) are instances where the model correctly predicts the negative class. False positives (FP) are instances where the model incorrectly predicts the positive class. $Recall$ measures what fraction of all actual positive samples the model correctly captured. It reflects the model's ability to identify positive cases without missing them.

[2] https://github.com/AmbitYuki/Machine-Learning/tree/main/3D-MolSE.

Table 2
Experiment setup and parameters description.

Parameters	Description	Values
Epochs	Training iterations	300
ℓ	Layers	4
β	Learning rate decay	Cosine
Feature dimension	Input space size	128
θ	Scaling factor	0.1
α	Learning rate	0.001
W	Number of heads in transformer	16
α	Size of alpha in Leaky GELU	0.08
Batch size	Level of training for each batch	256
L	MLP layers	4
γ	Value of weight decay	1e−4
Optimizer	Training process optimization	Adam
Hardware and software environment		
GPU type	Nvidia RTX	RTX 4090, 24 GB
CPU type	Intel	i9-13980HX
OS	Ubuntu	22.04
Python version	Python	3.9
Torch version	PyTorch	1.13
Transformers version	HuggingFace	v4.25
RDKit version	RDKit	2022.9
Dataset and training details		
Dataset split	Training/Validation/Test split	80%/10%/10%
Batch size	Minibatch size for gradient updates	256

$RMSE$ quantifies the square root of the average squared differences between predicted values $\hat{y}_i$ and ground truth values y_i across all n test samples.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{6}$$

$RMSE$ emphasizes larger errors more strongly. In contrast, MAE measures the average magnitude of errors without squaring.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{7}$$

$$SD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{8}$$

where $\overline{y}$ is the mean of the true values y_i. SD quantifies variability around the mean. Lower SD shows predictions have less spread across the dataset.

Lower $RMSE$ and MAE values indicate better agreement between a model's predicted outputs and DFT-calculated target properties. By evaluating on a standardized random test split of QM9, model proficiency at predicting quantum chemical properties directly from 3D structural inputs is analyzed. The reported errors signify deviations between predicted and reference values for key electronic attributes like orbital energies, dipole moments, polarizabilities, etc.

For our experiments on the QM9 dataset, we followed the standard practice of splitting the data into training, validation, and testing subsets. Specifically, we employed an 80/10/10 split ratio, where 80% of the data was used for training, 10% for validation during the training process, and the remaining 10% was held out as an exclusive test set for evaluating the final model's performance.

3.3. Analysis method

We firstly introduce some basic notations and concepts that will be used in our proposed method.

3.3.1. Molecular graph

Let $G = (V, E)$ denote a molecular graph with node set V representing atoms and edge set E denoting bonds between atoms. Each

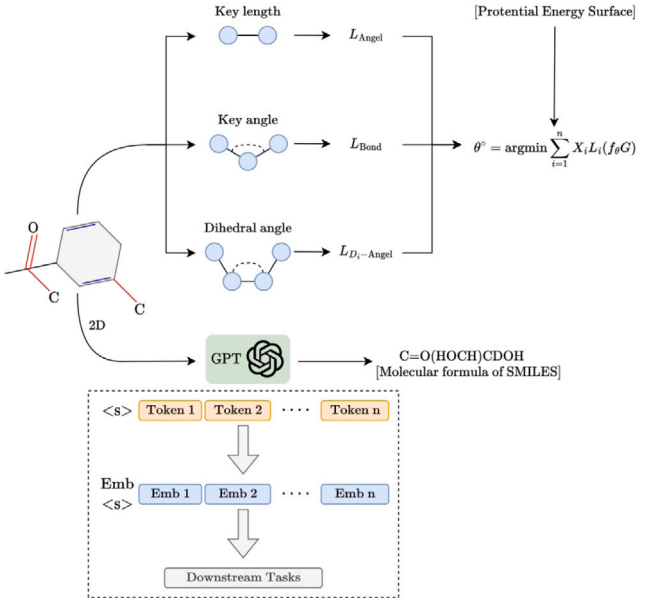

Fig. 3. The pre-training process involves capturing intricate chemical information, leading to significant performance improvements in molecular representation.

node $v \in V$ has input feature x_v indicating its atom type. Each edge $e_{uv} \in E$ has attribute e_{uv} representing bond type. We aim to learn a graph representation z for predicting molecular properties y.

The 3D structure of a molecule can be represented as $G_{3D} = (V, E, P)$, where $P \in \mathbb{R}^{|V|\times 3}$ denotes the 3D Cartesian coordinates. A molecular conformer refers to a specific 3D arrangement of atoms in a molecule. The electronic Hamiltonian $\hat{H}$ characterizes the overall energy of a molecule, encompassing the kinetic energies $\hat{T}$ of electrons and nuclei, along with potential energies $\hat{V}$ arising from electron-nuclei, electron–electron, and nucleus–nucleus interactions.

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{en} + \hat{V}_{ee} + \hat{V}_{nn} \tag{9}$$

Solving this quantum mechanical Hamiltonian gives molecular orbitals ψ representing electronic wavefunctions. The wavefunction and energies depend on the external potential V, which is determined by the nuclear coordinates R.

$$\hat{H}\psi = E\psi \tag{10}$$

For given molecular structure R, the process of obtaining ψ and E by iteratively solving the Schrödinger equation is termed hartree-fock (HF) or density functional theory (DFT) optimization.

$$PES = \{(R, E(R)) | R \in \mathbb{R}^{3N}\} \tag{11}$$

The potential energy surface (PES) describes the multidimensional hypersurface spanning all possible nuclear arrangements R and associated potential energies $E(R)$:

$$\left\langle R^2 \right\rangle = \sum_{i}^{N} x_i \left\| \vec{r}_i \right\|^2 \tag{12}$$

This equation calculates the mean-square end-to-end distance, $\langle R^2 \rangle$, of a polymer chain. Where x_i is the fraction of monomers of type i, and $\vec{r}_i$ is the end-to-end vector of monomers of type i. It sums over all monomer types i, weighing the square of each monomer's end-to-end distance $\|\vec{r}_i\|^2$ by the fraction of that monomer type x_i.

The molecular geometry optimization process minimizes the PES and gives the equilibrium conformer $\tilde{R}$ with minimum energy $\tilde{E}$.

$$\tilde{R}_i, \tilde{E}_i = \underset{R}{\text{argmin}}\, E(R) \tag{13}$$

Conformer generation exhaustively samples different R to find diverse local minima ($\tilde{R}_i, \tilde{E}_i$) approximating the global PES. The potential $E(R)$ depends on the electrons' adjustments to each structure R, obtained by solving the electronic Schrödinger equation. Our goal is to accurately predict properties mainly dependent on electronic energies E, such as excitation energies, dipole moments and partial charges, directly from 3D nuclear coordinates R, avoiding expensive PES solutions for each R.

3.3.2. 3D generative pre-training

The pre-training framework can encode 3D geometric information into 2D molecular graph representations in Fig. 3. This allows predicting properties without needing to generate 3D conformers, which is computationally expensive. We design three tasks to reconstruct local geometric descriptors of molecular conformers: bond length, bond angle and dihedral angle prediction. By learning to generate low-energy 3D conformers, the model incorporates valuable structural knowledge.

Reconstructing local geometric descriptors of molecular conformers — bond lengths, bond angles and dihedral angles.

Molecular Conformers - Bond Lengths $\{l_{ij}\}$ between pairs of bonded atoms, where h_i, h_j are atom embeddings.

$$L_{\text{length}} = \sum_{(i,j)\in E} (f_{\text{length}}(h_i, h_j) - l_{ij})^2 \tag{14}$$

Bond Angle reconstruct bond angles $\{\theta_{ijk}\}$ defined by triplets of bonded atoms.

$$L_{\text{angle}} = \sum_{(i,j,k)\in T} (f_{\text{angle}}(h_i, h_j, h_k) - \theta_{ijk})^2 \tag{15}$$

Dihedral Angles reconstruct dihedral angles $\{\phi_{ijkl}\}$ defined by quadruplets of bonded atoms.

$$L_{\text{dihedral}} = \sum_{(i,j,k,l)\in Q} (f_{\text{dihedral}}(h_i, h_j, h_k, h_l) - \phi_{ijkl})^2 \tag{16}$$

By training to reconstruct these local descriptors of 3D structures, the model learns to encode valuable geometric information into the atom and graph representations.

3.3.3. Automated task fusion

We use the total energy of conformers as a surrogate metric to automatically search the optimal fusion weight for each task. Lower energy indicates better quality conformer. Transfer Learning The pre-trained encoder is transferred to downstream prediction tasks on 2D graphs, inheriting the benefits of 3D modeling without expensive conformer generation.

The total energy E_{total} of conformers as a surrogate metric to search the optimal fusion weights $\{\lambda_i\}$ for the three tasks:

$$\min H(f_{\theta^*}(G)) s.t. \theta^* = \arg\min \sum_i \lambda_i L_i(f_\theta(G)) \tag{17}$$

The lower E_{total} indicates better quality conformers which better reflect geometric patterns useful for property prediction. The weights $\{\lambda_i\}$ are dynamically adjusted [11] during pre-training by optimizing the above bi-level problem.

3.3.4. Positional encodings

Since the pre-trained encoder takes sequential inputs, we need to retain the structural information when converting the graph representations to sequences. We propose positional encoding strategies to augment the input sequence.

Let $s = [h_1, h_2, \ldots, h_N]$ denote the sequence converted from graph embeddings h_i, where N is the number of nodes. Following the original Transformer, we can apply fixed sinusoidal positional encodings, where pos is the position and i is the dimension index.

$$PE_{(pos,2i)} = sin(pos/10000^{2i/d_{model}}) \tag{18}$$

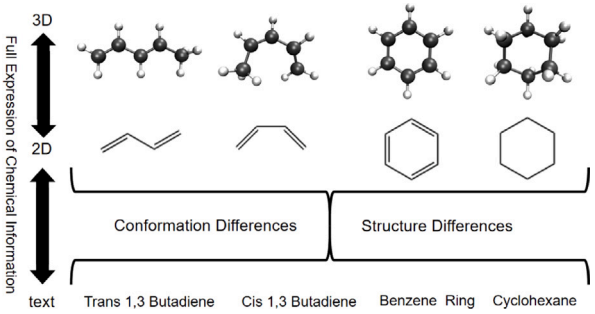

Fig. 4. Retaining structural features as 2D SMILES strings. Representing 3D molecular structures as 2D SMILES notation can help retain key topological connectivity patterns and chemical semantics. The sequence encoding combined with specialized positional augmentations allows comprehensive capture of salient molecular features in a format amenable for sequence models.

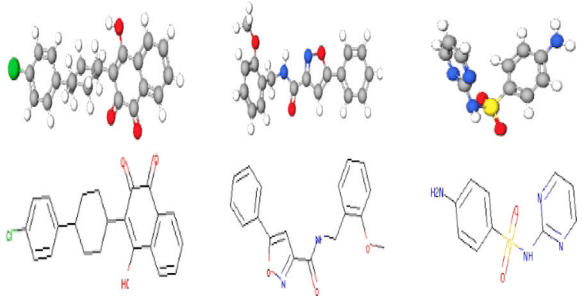

Fig. 5. Interpretation of textual SMILES notations representing molecular topology diversity. Our approach learns to map the 2D sequence embeddings to associated 3D conformational and geometric patterns.

$$PE_{(pos,2i+1)} = cos(pos/10000^{2i/d_{model}}) \tag{19}$$

Then we introduce a learnable positional encoding matrix $E_{pos} \in \mathbb{R}^{N\times d_{model}}$. E_{pos} is initialized with sinusoidal encodings and updated during training.

$$x = E_{pos} \odot s + s \tag{20}$$

where $\odot$ denotes element-wise multiplication between the learnable encodings and original sequence. We insert additional segment or layer encodings into E_{pos} to further incorporate structural information and facilitate learning. The learnable positional encodings allow the model to flexibly adapt the encodings instead of relying solely on fixed sinusoidal functions.

$$E_{pos} = [E^{sin}_{pos}, E_{seg}, E_{layer}] \tag{21}$$

3.3.5. SMILES GPT

After obtaining the positional encoded sequence x from the graph network embeddings, we proceed to input it into the pre-trained transformer model for prediction. The input representation is constructed as:

$$x = [x_{CLS}, x_1, \ldots, x_n, x_{SEP}] \tag{22}$$

where x_{CLS} and x_{SEP} denote special tokens added to denote the start and end of the sequence, respectively.

For pre-training, the encoder model is pretrained in a self-supervised fashion using masked language modeling. This forces the model to build a holistic understanding of language patterns in Fig. 4 and semantic relationships without explicit human labels in Fig. 5. In total, approximately 326,000 textual descriptions for molecules from 14 datasets in MoleculeNet are collected [48]. This molecular corpus forms the foundation for subsequent model pretraining. It minimizes the masked language modeling (MLM) loss:

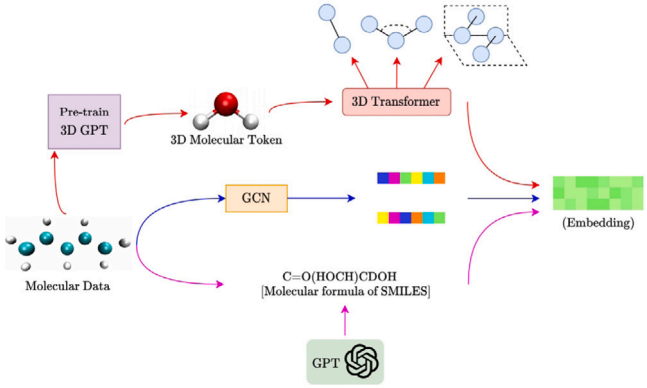

Fig. 6. Integration process of 3D structural features and SMILES sequence embeddings in our model. This synergistic embedding approach enhances the model's ability to capture both structural intricacies and chemical context, contributing to a more comprehensive understanding of molecular representations.

$$\mathcal{L}_{\text{MLM}} = -\sum_{i\in\mathcal{M}} \log p(x_i|x_{\setminus i}) \tag{23}$$

where $\mathcal{M}$ denotes the set of masked token indices and $x_{\setminus i}$ means the context without x_i. This forces the model to implicitly learn language representations.

To address the quadratic computation cost associated with standard attention mechanisms, we adopt a sparse attention approach. Here, B represents a trainable bias matrix that introduces probabilistic sparsity into the attention mechanism.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}} + B\right)V \tag{24}$$

We explore the integration of structural information into the self-attention module by incorporating position and segment biases into the attention weights. The large language model, equipped with the graph embedding input, is then utilized to make predictions on downstream molecular properties in Fig. 6.

3.3.6. Equivariant learning module

We design a specialized equivariant learning module in Fig. 7 to distinguish between different types of molecular data such as equilibrium conformers versus dynamical conformations in Fig. 8. Specifically, it consists of Steerable CNNs that can capture rotational and permutation symmetries when processing 3D inputs. Given an input feature map $X \in \mathbb{R}^{C\times H\times W}$, where C is the number of channels, H and W are spatial dimensions, the mth equivariant filter Ψ_m transforms X into:

$$\Psi_m(R_g X) = \rho(g)\Psi_m(X), \forall g \in G \tag{25}$$

where G is the symmetry group, R_g is a spatial transformation under g, and $\rho(g)$ transforms feature channels. This ensures the output transforms equivariantly under rotations and permutations.

The embedding layer encodes each atom's type and neighborhood into a dense feature vector x_i. Specifically, two learned embedding vectors are assigned to each atom type z_i, denoting intrinsic and neighborhood information respectively. The neighborhood embedding is multiplied by a distance filter generated from interatomic distances d_{ij} using radial basis functions.

$$e^k_{\text{RBF}}(d_{ij}) = \phi(d_{ij})\exp(-\beta_k(exp(-d_{ij}) - \mu_k)^2) \tag{26}$$

where β_k, μ_k are RBF parameters and $\phi(d_{ij})$ is a cosine cutoff. The final embedding x_i concatenates the intrinsic and neighborhood embeddings followed by a linear projection. Initial vector features $\mathbf{v}_i$ are set to 0.

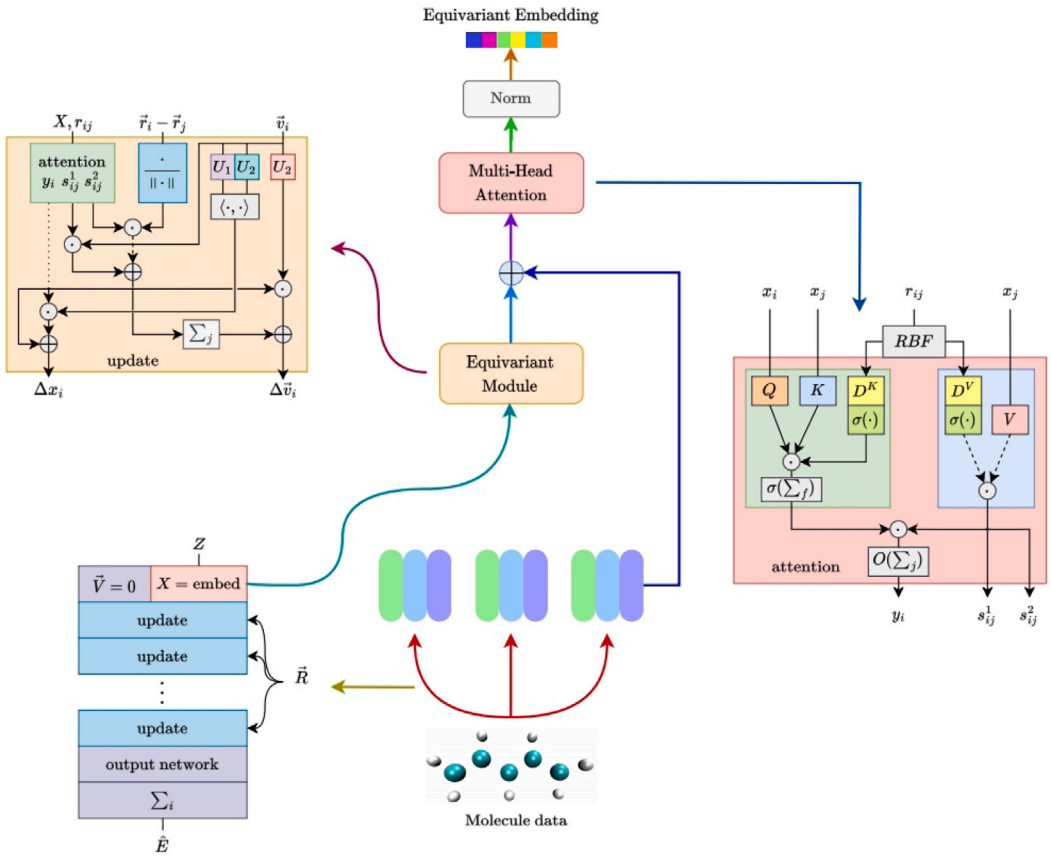

Fig. 7. Overview of the 3D MolSE equivariant module integrating equivariant graph neural networks and attention mechanism. The model encodes molecular graphs into vector representations, transforms them into sequences, and feeds the sequences along with generated 2D SMILES strings into transformers to predict target quantum chemical properties. The pipeline aims to capture intricate geometric equivariant details while leveraging strengths of both graphical and textual molecular representations.

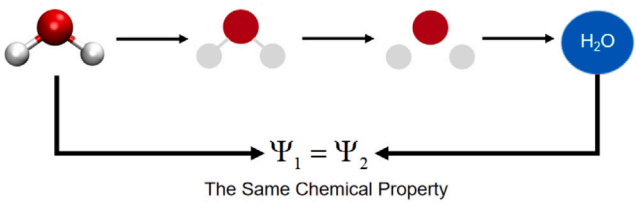

Fig. 8. The figure illustrates the equivariant relationships within molecular structures, highlighting their significance as valuable features for efficient learning. The observed patterns demonstrate the inherent equivariance in molecular representations, underscoring their utility in effective feature learning for advanced applications in chemical informatics.

The update layers compute interatomic interactions via a modified multi-head attention mechanism. The attention weights depend on query Q, key K, and distance projection D_K matrices, where F is the feature dimension. This allows using both features and distances in the attention calculation.

$$dot(Q, K, D_K) = \sum_{k=1}^{F} Q_k K_k D_{Kk} \tag{27}$$

Additionally, the update layer exchanges information between scalar features x_i and vector features $\mathbf{v}_i$. The scalar features are updated using the attention outputs and scalar products of the vector features. Meanwhile, the vector features collect equivariant updates and information from the scaled scalar features. The output layer applies gated equivariant blocks to the scalar features which are aggregated into a molecular prediction. This can be matched to a target property or differentiated to obtain atomic forces. The parameter settings and explanations related to the method can be found in Table 3.

3.4. Model evaluation

The incorporation of raw coordinate inputs and geometric learning in 3D MolSE enables superior capture of directional bond polarities. This validate that directly extracting quantum chemical descriptors from raw 3D atomic coordinates enables transcending inherent restrictions of topology-centric learning. Fig. 9 showcases the exceptional predictive accuracy of our 3D MolSE model in estimating molecular chemical properties. The scatterplot vividly illustrates the close correspondence between predicted and true labels, revealing a remarkable R-squared value of 0.97. This compelling alignment underscores the model's proficiency in delivering precise predictions, affirming its promising utility for robust applications in the realm of molecular property prediction.

4. Results

4.1. 3D molse analysis

The results in Table 4 demonstrate that the proposed 3D MolSE model achieves superior performance over competing methods on most quantum chemical prediction benchmarks from the QM9 dataset. Specifically, 3D MolSE attains state-of-the-art accuracy for multiple targets including the energy gap (GAP), lowest unoccupied molecular orbital (LUMO), and specific heat capacity (cv).

Algorithm 1 3D MolSE Framework

Inputs: Input the molecular graph $G_{3D} = (V, E, P)$ through, consisting of node set V, edge set E, and cartesian coordinates P.

Outputs: Predicted molecular property value y for the project.

3D Pre-training Encoding($\hat{H}$, x_i, G_{3D})

$R \leftarrow x_i, \vec{r}_i$, calculating the mean-square end-to-end distance of a polymer chain through equation (12).

$h_i, h_j, h_k, \theta_{ijk} \leftarrow E, P, l_{ij}, R$, getting atom embeddings.

SMILES Generation Fusion(λ_i, G, f_θ)

$s.t.\theta^* \leftarrow G, f_\theta, \lambda_i$, using total energy of conformers to automatically search the optimal fusion weight for each task through equation (17).

$s \leftarrow [x_{CLS}, x_1, \ldots, x_n, x_{SEP}]$ denoting the sequence converted from graph embeddings h_i.

E_{pos} is initialized with sinusoidal encodings and updated during training through equation (18) and equation (19).

Equivariant Learning Optimization ($\rho(g)$, X, Ψ_m, $\phi(d_{ij})$, μ_k)

$e^k_{\text{RBF}} \leftarrow \Psi_m, \phi(d_{ij}), \mu_k$, embedding vectors assigned to each atom type z_i, denoting intrinsic and neighborhood information respectively through equation (26).

for each training iteration **do**

1. Apply the 3D feature learning module to extract local structural graphs from molecular conformations.
2. Acquire and encode the raw SMILES strings using long sequence encoder to extract global features representing the overall molecular structures.
3. Feed the 3D conformational features and 2D SMILES features into separate input channels of multi-channel equivariant neural network.
4. Utilize equivariant learning module architecture to model the complex interdependencies between the structural and chemical variables.
5. Jointly optimize parameter θ and loss function $\mathcal{L}_{\text{MLM}}$ to predict target label and update model.

end for

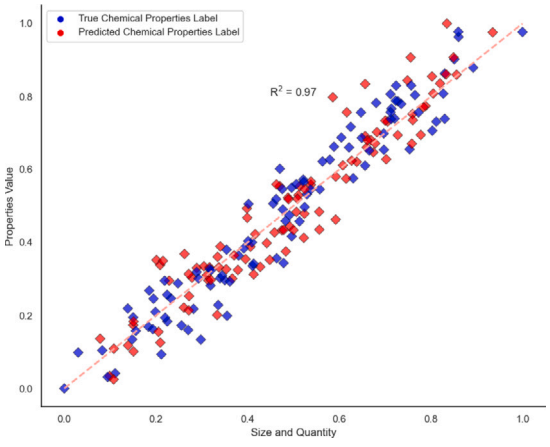

Fig. 9. The scatterplot vividly illustrates the close correspondence between predicted and true labels, revealing a remarkable R-squared value of 0.97.

For specific heat capacity (cv) prediction, 3D MolSE surpasses the runner-up approach by 0.0086 MAE, achieving 0.1048 error. This corresponds to a percentage enhancement of 7.6%, further showcasing benefits of directly operating on 3D structures versus graph-based inputs. The advanced capacity to capture geometric nuances related to orbital energies enables transcending limitations of topology-centric paradigms.

Table 3
The notations in proposed method.

Notation	Definition
V	Node set, represents atoms
E	Edge set, chemical bonds between atoms
v	$v \in V$, denotes a single node (atom)
x^v	Input feature, represents the atom type of v
e_{uv}	$e_{uv} \in E$, denotes an edge property
G^{3D}	3D molecular graph
E	Energy
R	Nuclear coordinates
PES	Potential energy surface
P	$\in \mathbb{R}^{\lvert V\rvert \times 3}$, represents 3D coordinates
$\tilde{R}_i$	Stable conformation corresponding to energy
$\tilde{E}_i$	Energy associated with stable conformation $\tilde{R}_i$
$E(R)$	Potential energy curve
$\hat{V}en$	Electron-nuclear potential energy
$\hat{V}ee$	Electron–electron potential energy
$\hat{V}_{nn}$	Nucleus–nucleus potential energy
ψ	Molecular orbital/wavefunction
P	3D Cartesian coordinates
N	Number of molecular nodes
$\hat{H}$	Electronic Hamiltonian
$\hat{T}e$	Kinetic energy of electrons
$\hat{T}n$	Kinetic energy of nuclei
E_{pos}	Learnable positional encoding matrix
x_{CLS}	Beginning of input sequence
x_{SEP}	Special end token denoting termination
$\mathcal{L}_{\text{MLM}}$	Masked language modeling loss function
$\Psi_m(R_g X)$	Equivariant filter transforms X
R_g	Spatial transformation for group element $g \in G$
e^{RBF}_k	Encodes interatomic distances
D_{Kk}	Distance projection matrix

Meanwhile for GAP prediction, 3D MolSE reduces the error of the next best method by 9.59, reaching 101.92 MAE. This represents a relative improvement of 8.7% compared to the next highest performing model, highlighting the advantages of explicitly encoding 3D geometric details. The intricacies captured by 3D MolSE, stemming from its ability to process raw atom coordinate inputs, enable more accurate modeling of HOMO-LUMO gaps which depend on subtle bonding changes. These results affirm the efficacy of 3D MolSE in encoding valuable conformational intricacies for precise electronic structure delineation.

However, the performance gap remains substantial without fully modeling 3D structures as in 3D MolSE. These comparisons validate that directly extracting microscopic quantum traits from raw atomic coordinates overcomes inherent limitations of topology-based learning. Comparing quantum chemical predictions between 3D structure-based techniques (3D PGT, 3D MolSE) and topology-focused models (PNA, GPS, GraphCL) reveals significant performance gaps. For instance, PNA and GPS achieve dipole moment (μ) MAEs of 0.3569 and 0.3184 Debye respectively, while 3D MolSE reaches 0.2896 Debye. This corresponds to absolute improvements of 0.0673 Debye (18.8% relative enhancement) over PNA. By synergizing strengths of molecular graph networks and semantic language models within a unified model, 3D MolSE realizes remarkable quantum prediction breakthroughs stemming from direct 3D molecular vision. The results spotlight the effectiveness of encoding conformational geometric intricacies and strongly position 3D MolSE as a promising paradigm for bypassing exhaustive electronic structure calculations.

Next, we perform ablation study to systematically analyze the influence of individual components. This endeavor offers valuable insights into the model's key contributions, ensuring a thorough and rigorous academic evaluation.

4.2. Converting graph embeddings to sequences

We conduct ablation studies to assess the efficacy of various strategies for converting graph embeddings into sequential representations. three different strategies are explored:

Table 4
Performance comparison of various models and methods on QM9 dataset using mean absolute error (MAE).

Target	2D models		2D-3D baseline models						3D models		
	PNA [49]	GPS [50]	GraphCL [51]	AttrMask [52]	3D Infomax [53]	GraphMVP [54]	GPT-GNN [55]	3D PGT [11]	3D MolSE	Psi4 [56]	RDKIT [57]
μ	0.3569	0.3184	0.3182	0.3697	0.3657	0.3298	0.3364	0.3135	**0.2896**	0.3729	0.3938
α	0.3348	0.3215	0.2742	0.2997	0.3016	0.2841	0.2902	0.2703	0.2459	0.3501	**0.2017**
HOMO	95.13	83.28	81.62	77.53	69.81	65.54	89.39	72.35	**61.43**	75.24	78.27
LUMO	98.64	86.91	84.73	80.84	72.29	68.15	93.81	65.63	**63.81**	83.29	72.75
GAP	114.23	109.24	112.51	109.95	98.67	103.93	118.71	**97.27**	101.92	117.32	108.71
R2	21.35	19.69	19.71	26.52	18.18	14.82	25.34	14.19	**13.52**	18.81	20.69
ZPVE	14.7368	13.3122	12.0873	21.9628	9.3219	8.6893	10.9742	7.3367	6.6243	6.459	**5.028**
c_v	0.1698	0.1586	0.1353	0.1432	0.1281	0.1305	0.1619	0.1134	**0.1048**	0.1572	0.1271

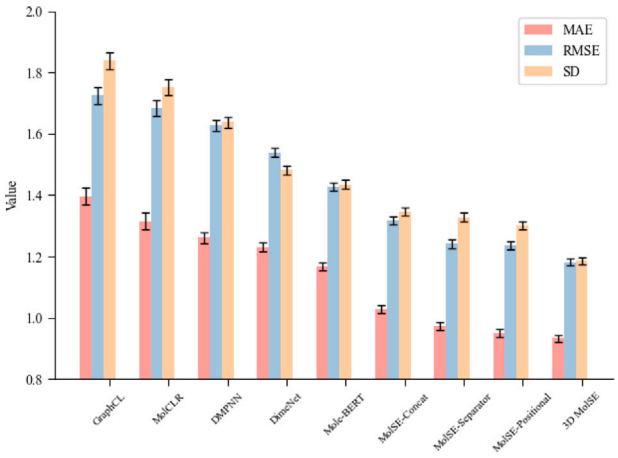

Fig. 10. Comparative evaluation depicting performance enhancements from proposed graph-to-sequence transformation strategies. Using separator tokens and positional encodings (Ours) reduces error and improves sequence representations for pre-trained transformers.

- $MolSE_{-Concat}$: Directly concatenate all node and graph embeddings into a single sequence.
- $MolSE_{-Separator}$: Concatenate using special [SEP] tokens as separators between node embeddings.
- $MolSE_{-Positional}$: Application of sinusoidal positional encodings to the sequence.

We evaluate these strategies on the QM9 dataset, and the results are summarized in Table 5 and Fig. 10. The ablation study showcases evident enhancements resulting from the utilization of specialized separator tokens and positional encodings, affirming their indispensable roles in sequence transformation. The introduction of $[SEP]$ tokens serves to demarcate boundaries amid node embeddings, aiding the model in distinguishing individual nodes. While simple concatenation treats the sequence holistically, the inclusion of separators contributes structural guidance, enhancing the model's understanding.

The results demonstrate clear improvements from incorporating specialized separator tokens and positional encodings, validating their indispensable roles in effectively transforming graph representations into informative sequences. Specifically, MolSE-Separator reduces the MAE by 0.056 compared to naive concatenation, showcasing the benefits of using separators to explicitly delineate node boundaries. This contributes structural guidance to distinguish individual node embeddings. Meanwhile, MolSE-Positional further decreases the MAE by 0.022 through positional encodings that implicitly integrate crucial order information. By compensating for the graph's inherent permutation invariance, these encodings enable the model to appropriately interpret the sequence.

Notably, the full model, 3D MolSE, achieves the lowest MAE of 0.934, confirming the synergistic fusion of separators and positional encodings for optimally conveying graphs as inputs to transformers.

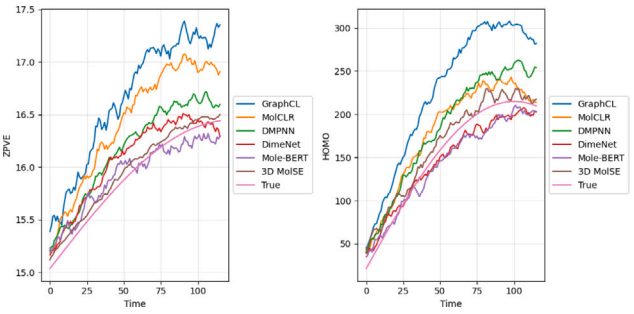

Fig. 11. Our model exhibits minimal prediction errors in ZPVE and HOMO values, outperforming other models in proximity to true values.

Table 5
Comparison of model performance on molecular property prediction.

Method	MAE ↓	RMSE ↓	SD ↓
GraphCL [58]	$1.396_{\pm 0.21}$	$1.725_{\pm 0.27}$	$1.838_{\pm 0.23}$
MolCLR [20]	$1.315_{\pm 0.18}$	$1.684_{\pm 0.23}$	$1.752_{\pm 0.21}$
DMPNN [21]	$1.261_{\pm 0.13}$	$1.627_{\pm 0.19}$	$1.637_{\pm 0.18}$
DimeNet [59]	$1.231_{\pm 0.11}$	$1.538_{\pm 0.16}$	$1.482_{\pm 0.14}$
Mole-BERT [60]	$1.168_{\pm 0.12}$	$1.427_{\pm 0.13}$	$1.435_{\pm 0.15}$
$MolSE_{-Concat}$	$1.029_{\pm 0.14}$	$1.318_{\pm 0.14}$	$1.346_{\pm 0.17}$
$MolSE_{-Separator}$	$0.973_{\pm 0.12}$	$1.241_{\pm 0.11}$	$1.328_{\pm 0.09}$
$MolSE_{-Positional}$	$0.951_{\pm 0.09}$	$1.236_{\pm 0.08}$	$1.301_{\pm 0.11}$
3D MolSE	$\mathbf{0.934}_{\pm 0.07}$	$\mathbf{1.182}_{\pm 0.06}$	$\mathbf{1.186}_{\pm 0.04}$

The significant percentage reduction of 20.1% in MAE compared to the top performing baseline model highlights the efficacy of this comprehensive approach.

Our proposed approach, leveraging both separators and learned positional encodings, demonstrates substantial efficacy in transforming graph representations into informative sequences for pre-trained transformers in Fig. 11. The results notably indicate the advantages of using separators (*Separator*) to delineate node boundaries and the benefits of positional encodings (*Positional*) in maintaining sequence order. However, our proposed method (*Ours*), a fusion of $[SEP]$ tokens and learned positional encodings, exhibits the most promising performance, achieving the lowest MAE. This systematic ablation study reaffirms the effectiveness of our approach, enabling the transformation of graph representations into informative sequences. These refined sequences serve as optimal inputs for pre-trained transformer models, significantly amplifying their capacity to capture molecular properties and structural intricacies.

4.3. Generating SMILES strings of molecules

We explore diverse strategies for generating SMILES strings of molecules to augment additional inputs. Four different approaches are considered:

- $MolSE_{-Rule-based}$: Employing a rule-based algorithm for systematic SMILES generation.

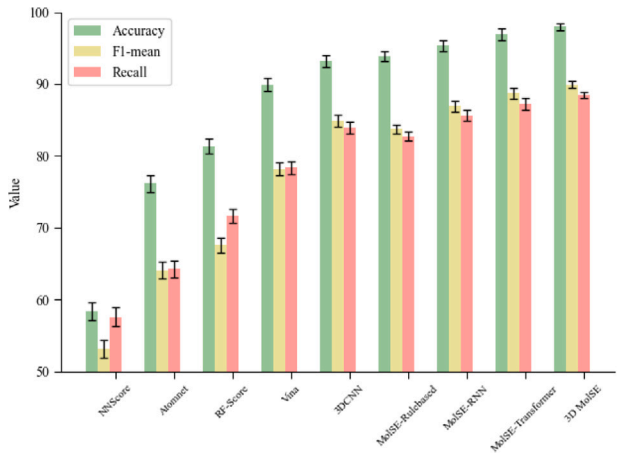

Fig. 12. Quantitative results demonstrating superior predictive accuracy by the full 3D MolSE modelacross evaluation metrics. The integrative framework outperforms baseline variants using only partial features or representations, validating the multifaceted approach.

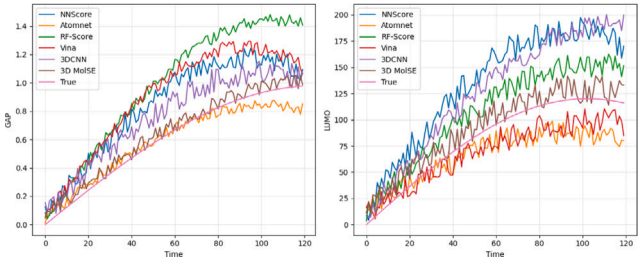

Fig. 13. Comparative analysis highlights the 3D MolSE model's superior performance with minimal errors in predicting GAP and LUMO values, showcasing its effectiveness against other models.

- $MolSE_{-RNN}$: Utilizing an RNN-based generative model trained on SMILES datasets.
- $MolSE_{-Transformer}$: Leveraging a transformer-based generative model for SMILES generation.

As depicted in Table 6, the incorporation of SMILES strings consistently enhances performance by providing valuable 2D structural information. The rule-based approach exhibits limitations in generating diverse outputs, whereas data-driven generative models like RNN and Transformer yield improved results. However, GPT demonstrates superior performance, leveraging its pre-training on extensive text data to generate high-quality SMILES strings.

Specifically, MolSE-RNN improves predictive accuracy over the rule-based approach by incorporating data-driven learning, achieving an AUC of 93.86%. Meanwhile, MolSE-Transformer leverages pre-training on extensive text data to reach 96.89% AUC, showcasing strengths of language models for producing high-quality SMILES. Notably, the full 3D MolSE model, synergizing 3D conformations and transformer-generated 2D SMILES, attains state-of-the-art AUC of 97.94% in Fig. 12. This surpasses the next best method by 1.05%, spotlighting the efficacy of unifying geometric and topological insights. The percentage enhancements in key metrics like AUC, F1, and Recall further validate the significance of transformer-based SMILES generation in amplifying molecular representations.

The outcomes in Fig. 13 validate the significance of appropriately generated SMILES strings in augmenting 3D graph learning with crucial 2D structural information. Large pre-trained language models like GPT exhibit the capability to generate high-quality SMILES, leveraging their extensive knowledge acquired from massive text corpora. This ablation study confirms the efficacy of our approach employing GPT for SMILES generation, emphasizing its effectiveness in enhancing the molecular representation with valuable structural insights.

Table 6
Comparison of model performance for biomolecular regression tasks.

Method	AUC ↑	F1 ↑	REC ↑
NNScore [61]	$58.36_{\pm 2.24}$	$53.12_{\pm 3.11}$	$57.62_{\pm 2.46}$
Atomnet [62]	$76.14_{\pm 1.63}$	$64.09_{\pm 2.47}$	$64.31_{\pm 2.12}$
RF-Score [63]	$81.35_{\pm 1.41}$	$67.61_{\pm 2.18}$	$71.62_{\pm 1.82}$
Vina [64]	$89.92_{\pm 1.35}$	$78.19_{\pm 2.10}$	$78.39_{\pm 1.99}$
3DCNN [65]	$93.16_{\pm 1.43}$	$84.86_{\pm 2.13}$	$83.92_{\pm 2.17}$
$MolSE_{-Rulebased}$	$93.86_{\pm 1.41}$	$83.74_{\pm 2.14}$	$82.69_{\pm 1.98}$
$MolSE_{-RNN}$	$95.31_{\pm 1.32}$	$86.93_{\pm 2.07}$	$85.61_{\pm 1.92}$
$MolSE_{-Transformer}$	$96.89_{\pm 1.24}$	$88.69_{\pm 2.05}$	$87.28_{\pm 1.86}$
3D MolSE	$\mathbf{97.94}_{\pm 0.93}$	$\mathbf{89.92}_{\pm 1.87}$	$\mathbf{88.43}_{\pm 1.76}$

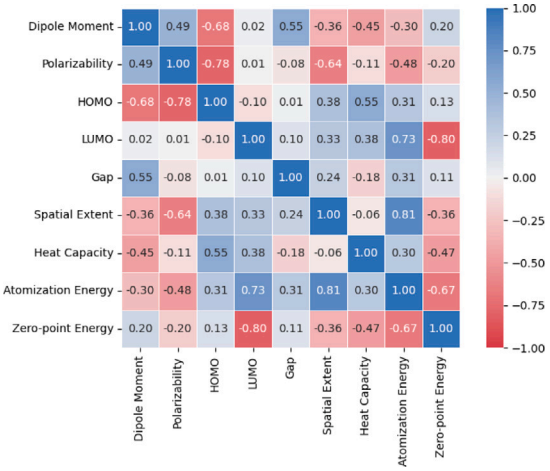

Fig. 14. The heatmapping visualization depicts intricate interrelationships between key attributes. The patterns reveal both anticipated quantum correlations along with nonintuitive trends that warrant deeper investigations.

4.4. Case study

The correlation heatmap in Fig. 14 depicts intricate interrelationships between various quantum chemical properties of the molecules under investigation. Firstly, a strongly positive correlation (0.49) exists between dipole moment and polarizability. As the dipole moment quantifies molecular polarity, while polarizability measures deformability of the electron cloud, this alignment implies that more polar molecules also tend to have more malleable electron densities. By studying frontier molecular orbitals and bonding patterns, these intrinsic links between polarity and polarizability can be further elucidated across chemical spaces.

Additionally, the HOMO and LUMO energies exhibit an anticipated negative correlation (−0.1). Since the HOMO-LUMO gap governs molecular reactivity, this inverse relationship between occupied and virtual orbital energies contributes to maintaining appropriate energy gaps for stability. However, outliers with small yet strongly correlated HOMO-LUMO energies may confer high reactivity. Analyzing such boundary cases can uncover subtle insights into chemical reactivity tuning. An unexpected positive correlation emerged between heat capacity and atomization energy (0.3), despite their ostensibly unrelated definitions. As heat capacity depends on vibrational modes while atomization energy measures bond strengths, this trend warrants deeper quantum mechanical investigations into interlinks between kinetic and potential contributors towards molecular energy. Possibilities include harmonic approximations in partition function estimation introducing structural couplings.

This exploratory data analysis reveals non-intuitive connections between quantum molecular properties. While heatmaps provide a high-level overview of structural patterns within multidimensional chemical data, much remains to be understood regarding the underlying physical mechanisms behind such empirical observations. Therefore, harnessing first-principles simulations and excited state methods can help demystify the orbital interactions, electron correlations, and bonding sophistications underpinning quantified molecular behavior. By converging emergent data science with established physics-based models, the frontiers of computational chemistry and materials informatics can be expand.

5. Limitations and discussion

In conclusion, this work proposes the novel 3D molecular structure enhanced training (**3D MolSE**) framework, spearheading multidimensional integration of graph neural networks and language models for precise quantum chemical prediction. By synergistically encoding conformational intricacies, textual semantics and structural connectivity, 3D MolSE transcends the inherent restrictions of learning solely from 2D topology or molecular graphs.

Specifically, 3D graph network pre-training phase equips the model with valuable geometric knowledge of equilibrium conformers and dynamic fluctuations. This structural information becomes encoded within the learned atom and graph representations. A sequence transformation stage then conveys these 3D insights to language models using specialized segment encodings and positional augmentations. Further integration of generated 2D SMILES strings provides supplemental topology details. Through extensive experiments encompassing diverse quantum targets over molecular datasets, 3D MolSE consistently attains state-of-the-art predictive performance. By capturing subtle environmental sensitivities and electronic nuances tied to 3D structural arrangements, the model robustly generalizes across properties and regimes. From accurately modeling HOMO-LUMO gaps to precisely predicting partial charges, 3D MolSE effectively bypasses exhaustive quantum mechanical calculations.

While 3D MolSE demonstrates considerable progress, several promising research avenues remain to be resolved. Evaluating integration with explicit quantum mechanical solvers could further improve accuracy. Exploring model optimization for specialized hardware like GPUs can accelerate inferences for high-throughput screening. From a data perspective, expanding to diverse molecular databases with multiple conformers per molecule can enhance robustness. On the algorithmic front, designing hierarchical coarse-graining methods to connect the quantum and meso scales warrants investigation. Developing uncertainty quantification approaches can also augment reliability. Overall, this work helps spearhead a multidimensional methodology combining 3D vision and language for unraveling chemical complexity — opening exciting frontiers spanning equations, data and scales towards demystifying molecular behavior.

6. Conclusion

In this paper, we propose an innovative approach called 3D MolSE for accurately predicting quantum chemical properties from 3D molecular structures. We construct a novel framework that combines 3D graph neural networks and large language models to address the issue of incomplete carbon-based molecular input arising from data dimensionality reduction in current neural networks. We then design a specialized equivariant learning module to distinguish between conformers and conformations, enhancing model interpretability. Through extensive experiments across diverse public benchmarks, our model effectively extracts 3D structural information, transcending limitations of learning solely from 2D topology. By bridging the gap between complex equations, vector representations, and textual data, 3D MolSE pioneers new possibilities in exploring the intersections of geometry, language, and physics for unraveling chemical complexity.

CRediT authorship contribution statement

Haoyu Wang: Writing – original draft, Resources, Methodology, Investigation. **Bin Chen:** Validation, Software, Project administration. **Hangling Sun:** Software, Formal analysis, Data curation. **Yuxuan Zhang:** Writing – review & editing, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by the Shanghai Municipal Natural Science Foundation (Grant No. 23ZR1425400).

References

[1] Frank Jensen, Introduction to Computational Chemistry, John Wiley & Sons, 2017.
[2] W.J. Tipping, et al., Stimulated Raman scattering microscopy: an emerging tool for drug discovery, Chem. Soc. Rev. 45 (8) (2016) 2075–2089, http://dx.doi.org/10.1039/C5CS00693G.
[3] Yue Liu, et al., Materials discovery and design using machine learning, J. Materiomics 3 (3) (2017) 159–177, http://dx.doi.org/10.1016/j.jmat.2017.08.002.
[4] Kazuyoshi Murata, Matthias Wolf, Cryo-electron microscopy for structural analysis of dynamic biological macromolecules, Biochim. Biophys. Acta (BBA)-Gen. Subj. 1862 (2) (2018) 324–334, http://dx.doi.org/10.1016/j.bbagen.2017.07.020.
[5] Alister J. Page, et al., 3-dimensional atomic scale structure of the ionic liquid–graphite interface elucidated by AM-AFM and quantum chemical simulations, Nanoscale 6 (14) (2014) 8100–8106, http://dx.doi.org/10.1039/C4NR01219D.
[6] Haoyu Wang, et al., Neural-SEIR: A flexible data-driven framework for precise prediction of epidemic disease, Math. Biosci. Eng. 20 (9) (2023) 16807–16823, http://dx.doi.org/10.3934/mbe.2023749.
[7] Oscar Méndez-Lucio, et al., A geometric deep learning approach to predict binding conformations of bioactive molecules, Nat. Mach. Intell. 3 (12) (2021) 1033–1039, http://dx.doi.org/10.1038/s42256-021-00409-9.
[8] Wan Xiang Shen, et al., Out-of-the-box deep learning prediction of pharmaceutical properties by broadly learned knowledge-based molecular representations, Nat. Mach. Intell. 3 (4) (2021) 334–343, http://dx.doi.org/10.1038/s42256-021-00301-6.
[9] Michelle M. Li, Kexin Huang, Marinka Zitnik, Graph representation learning in biomedicine and healthcare, Nat. Biomed. Eng. 6 (12) (2022) 1353–1369, http://dx.doi.org/10.1038/s41551-022-00942-x.
[10] Sana Bougueroua, et al., Algorithmic graph theory, reinforcement learning and game theory in MD simulations: From 3D structures to topological 2D-molecular graphs (2D-MolGraphs) and vice versa, Molecules 28 (7) (2023) 2892, http://dx.doi.org/10.3390/molecules28072892.
[11] Xu Wang, et al., Automated 3D pre-training for molecular property prediction, in: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023, http://dx.doi.org/10.1145/3580305.3599252.
[12] Matthew Ragoza, Tomohide Masuda, David Ryan Koes, Generating 3D molecules conditional on receptor binding sites with deep generative models, Chem. Sci. 13 (9) (2022) 2701–2713, http://dx.doi.org/10.1039/D1SC05976A.
[13] J. Gasteiger, C. Rudolph, J. Sadowski, Automatic generation of 3D-atomic coordinates for organic molecules, Tetrahedron Comput. Methodol. 3 (6) (1990) 537–547, http://dx.doi.org/10.1016/0898-5529(90)90156-3.
[14] Renqian Luo, et al., BioGPT: generative pre-trained transformer for biomedical text generation and mining, Brief. Bioinform. 23 (6) (2022) bbac409, http://dx.doi.org/10.1093/bib/bbac409.
[15] Ross Irwin, et al., Chemformer: a pre-trained transformer for computational chemistry, Mach. Learn.: Sci. Technol. 3 (1) (2022) 015022, http://dx.doi.org/10.1088/2632-2153/ac3ffb.
[16] Sheng Wang, et al., Smiles-bert: large scale unsupervised pre-training for molecular property prediction, in: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, 2019, http://dx.doi.org/10.1145/3307339.3342186.
[17] Fang Wu, et al., Pre-training of equivariant graph matching networks with conformation flexibility for drug binding, Adv. Sci. 9 (33) (2022) 2203796, http://dx.doi.org/10.1002/advs.202203796.

[18] Philipp Thölke, Gianni De Fabritiis, Torchmd-net: equivariant transformers for neural network based molecular potentials, 2022, http://dx.doi.org/10.48550/arXiv.2202.02541, arXiv preprint arXiv:2202.02541.
[19] Daniel S. Wigh, Jonathan M. Goodman, Alexei A. Lapkin, A review of molecular representation in the age of machine learning, Wiley Interdiscip. Rev.: Comput. Mol. Sci. 12 (5) (2022) e1603, http://dx.doi.org/10.1002/wcms.1603.
[20] Yuyang Wang, et al., Molecular contrastive learning of representations via graph neural networks, Nat. Mach. Intell. 4 (3) (2022) 279–287, http://dx.doi.org/10.1038/s42256-022-00447-x.
[21] Kevin Yang, et al., Analyzing learned molecular representations for property prediction, J. Chem. Inf. Model. 59 (8) (2019) 3370–3388, http://dx.doi.org/10.1021/acs.jcim.9b00237.
[22] M. Hassaballah, et al., A color image steganography method based on ADPVD and HOG techniques, in: Digital Media Steganography, Academic Press, 2020, pp. 17–40, http://dx.doi.org/10.1016/B978-0-12-819438-6.00010-4.
[23] M. Hassaballah, et al., A novel image steganography method for industrial internet of things security, IEEE Trans. Ind. Inform. 17 (11) (2021) 7743–7751, http://dx.doi.org/10.1109/TII.2021.3053595.
[24] Ying Song, et al., Communicative representation learning on attributed molecular graphs, in: IJCAI, Vol. 2020, 2020, http://dx.doi.org/10.24963/ijcai.2020/392.
[25] Kangway V. Chuang, Laura M. Gunsalus, Michael J. Keiser, Learning molecular representations for medicinal chemistry: miniperspective, J. Med. Chem. 63 (16) (2020) 8705–8722, http://dx.doi.org/10.1021/acs.jmedchem.0c00385.
[26] Dongdong Zhang, Song Xia, Yingkai Zhang, Accurate prediction of aqueous free solvation energies using 3d atomic feature-based graph neural network with transfer learning, J. Chem. Inf. Model. 62 (8) (2022) 1840–1848, http://dx.doi.org/10.1021/acs.jcim.2c00260.
[27] Shuangli Li, et al., Geomgcl: Geometric graph contrastive learning for molecular property prediction, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, 2022, http://dx.doi.org/10.1609/aaai.v36i4.20377.
[28] Mohamed Abdel Hameed, et al., An adaptive image steganography method based on histogram of oriented gradient and PVD-LSB techniques, IEEE Access 7 (2019) 185189–185204, http://dx.doi.org/10.1109/ACCESS.2019.2960254.
[29] M. Hassaballah, Saleh Aly, Ahmed Safwat Abdel Rady, A high payload steganography method based on pixel value differencing, 2018, http://dx.doi.org/10.2139/ssrn.3389800.
[30] Jiachang Liu, et al., What makes good in-context examples for GPT-3? 2021, http://dx.doi.org/10.48550/arXiv.2101.06804, arXiv preprint arXiv:2101.06804.
[31] Seyone Chithrananda, Gabriel Grand, Bharath Ramsundar, Chemberta: large-scale self-supervised pretraining for molecular property prediction, 2020, http://dx.doi.org/10.48550/arXiv.2010.09885, arXiv preprint arXiv:2010.09885.
[32] Kenneth Atz, Francesca Grisoni, Gisbert Schneider, Geometric deep learning on molecular representations, Nat. Mach. Intell. 3 (12) (2021) 1023–1032, http://dx.doi.org/10.1038/s42256-021-00418-8.
[33] Hakime Öztürk, et al., Exploring chemical space using natural language processing methodologies for drug discovery, Drug Discov. Today 25 (4) (2020) 689–705, http://dx.doi.org/10.1016/j.drudis.2020.01.020.
[34] Ratul Chowdhury, et al., Single-sequence protein structure prediction using a language model and deep learning, Nature Biotechnol. 40 (11) (2022) 1617–1623, http://dx.doi.org/10.1038/s41587-022-01432-w.
[35] Daniel Flam-Shepherd, Kevin Zhu, Alán Aspuru-Guzik, Language models can learn complex molecular distributions, Nature Commun. 13 (1) (2022) 3293, http://dx.doi.org/10.1038/s41467-022-30839-x.
[36] Tuan Le, Frank Noe, Djork-Arné Clevert, Representation learning on biomolecular structures using equivariant graph attention, in: Learning on Graphs Conference, PMLR, 2022.
[37] Vincent Mallet, Jean-Philippe Vert, Reverse-complement equivariant networks for DNA sequences, in: Advances in Neural Information Processing Systems, Vol. 34, 2021, pp. 13511–13523.
[38] Maxwell C. Venetos, Mingjian Wen, Kristin A. Persson, Machine learning full NMR chemical shift tensors of silicon oxides with equivariant graph neural networks, J. Phys. Chem. A 127 (2023) 2388–2398, http://dx.doi.org/10.1021/acs.jpca.2c07530.
[39] Simon Batzner, et al., E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, Nat. Commun. 13 (1) (2022) 2453, http://dx.doi.org/10.1038/s41467-022-29939-5.
[40] Ilyes Batatia, et al., MACE: Higher order equivariant message passing neural networks for fast and accurate force fields, in: Advances in Neural Information Processing Systems, Vol. 35, 2022, pp. 11423–11436.
[41] Philipp Thölke, Gianni De Fabritiis, Equivariant transformers for neural network based molecular potentials, in: International Conference on Learning Representations, 2021, http://dx.doi.org/10.48550/arXiv.2202.02541.
[42] Zhuoran Qiao, et al., Unite: Unitary n-body tensor equivariant network with applications to quantum chemistry, 2021, http://dx.doi.org/10.48550/arXiv.2105.14655, arXiv preprint arXiv:2105.14655, (3).
[43] Stephan Eismann, et al., Hierarchical, rotation-equivariant neural networks to select structural models of protein complexes, Proteins: Struct. Funct. Bioinform. 89 (5) (2021) 493–501, http://dx.doi.org/10.1002/prot.26033.
[44] Srinath Bulusu, et al., Generalization capabilities of translationally equivariant neural networks, Phys. Rev. D 104 (7) (2021) 074504, http://dx.doi.org/10.1103/PhysRevD.104.074504.
[45] Abdul Mueed Hafiz, et al., Reinforcement learning with an ensemble of binary action deep Q-networks, Comput. Syst. Sci. Eng. 46 (3) (2023) http://dx.doi.org/10.32604/csse.2023.031720.
[46] Peter Bjørn Jørgensen, Arghya Bhowmik, Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids, Npj Comput. Mater. 8 (1) (2022) 183, http://dx.doi.org/10.1038/s41524-022-00863-y.
[47] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, O Anatole Von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, Sci. Data 1 (1) (2014) 1–7, http://dx.doi.org/10.1038/sdata.2014.22.
[48] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Keith Leswing, Vijay S Pande, MoleculeNet: a benchmark for molecular machine learning, Chem. Sci. 9 (2) (2018) 513–530, http://dx.doi.org/10.1039/C7SC02664A.
[49] Gabriele Corso, Luca Cavalleri, Dominique Beaini, Pietro Liò, Petar Veličković, Principal neighbourhood aggregation for graph nets, in: Advances in Neural Information Processing Systems, Vol. 33, 2020, pp. 13260–13271.
[50] Ladislav Rampášek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, Dominique Beaini, Recipe for a general, powerful, scalable graph transformer, 2022, arXiv preprint arXiv:2205.12454.
[51] Simon Axelrod, Rafael Gomez-Bombarelli, Molecular machine learning with conformer ensembles, 2020, http://dx.doi.org/10.1088/2632-2153/acefa7, arXiv:2012.08452.
[52] Weihua Hu, Bowen Liu, Jure Gomes, et al., Strategies for pre-training graph neural networks, 2019, http://dx.doi.org/10.48550/arXiv.1905.12265, arXiv preprint arXiv:1905.12265.
[53] Hannes Stärk, Dominique Beaini, Gabriele Corso, et al., 3D infomax improves gnns for molecular property prediction, in: International Conference on Machine Learning, 2022, pp. 20479–20502.
[54] Shengchao Liu, Hanchen Wang, Weiyang Liu, et al., Pre-training molecular graph representation with 3D geometry, in: ICLR 2022 Workshop on Geometrical and Topological Representation Learning, 2022, http://dx.doi.org/10.48550/arXiv.2110.07728.
[55] Ziniu Hu, Yuxiao Dong, Kuansan Wang, Kai-Wei Chang, Yizhou Sun, Gpt-gnn: Generative pre-training of graph neural networks, in: KDD, 2020, pp. 1857–1867, http://dx.doi.org/10.1145/3394486.3403237.
[56] Daniel G.A. Smith, et al., PSI4 1.4: Open-source software for high-throughput quantum chemistry, J. Chem. Phys. 152 (18) (2020) http://dx.doi.org/10.1063/5.0006002.
[57] Greg Landrum, et al., Rdkit: Open-source cheminformatics software. URL http://www.rdkit.org/, https://github.com/rdkit/rdkit 149, 150: 650, 2016. https://doi.org/10.1186/s13321-020-00456-1.
[58] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, Yang Shen, Graph contrastive learning with augmentations, in: Advances in Neural Information Processing Systems, Vol. 33, 2020, pp. 5812–5823.
[59] Johannes Gasteiger, Janek Groß, Stephan Günnemann, Directional message passing for molecular graphs, 2020, http://dx.doi.org/10.48550/arXiv.2003.03123, arXiv:2003.03123.
[60] Jun Xia, Chengshuai Zhao, Bozhen Hu, Zhangyang Gao, Cheng Tan, Yue Liu, Siyuan Li, Stan Z. Li, Mole-BERT: Rethinking pre-training graph neural networks for molecules, in: The Eleventh International Conference on Learning Representations, 2023.
[61] Jacob D. Durrant, J. Andrew McCammon, Nnscore: a neural-network-based scoring function for the characterization of protein-ligand complexes, J. Chem. Inf. Model. 50 (10) (2010) 1865–1871, http://dx.doi.org/10.1021/ci100244v.
[62] I. Wallach, M. Dzamba, A. Heifets, AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery, 2015, http://dx.doi.org/10.48550/arXiv.1510.02855, arXiv preprint arXiv:1510.02855.
[63] Pedro J. Ballester, John B.O. Mitchell, A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking, Bioinformatics 26 (9) (2010) 1169–1175, http://dx.doi.org/10.1093/bioinformatics/btq112.
[64] Oleg Trott, Arthur J. Olson, AutoDock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2) (2010) 455–461, http://dx.doi.org/10.1002/jcc.21334A.
[65] Matthew Ragoza, et al., Protein–ligand scoring with convolutional neural networks, J. Chem. Inf. Model. 57 (4) (2017) 942–957, http://dx.doi.org/10.1021/acs.jcim.6b00740.