Accurate and Generalizable Protein-Ligand Binding Affinity Prediction With Geometric Deep Learning

Krinos Li, Xianglu Xiao, Zijun Zhong, and Guang Yang, *Senior Member, IEEE*

Abstract*—*Goal: **Protein-ligand binding complexes are ubiquitous and essential to life. Protein-ligand binding affinity prediction (PLA) quantifies the binding strength between ligands and proteins, providing crucial insights for discovering and designing potential candidate ligands. While recent advances have been made in predicting protein-ligand complex structures, existing algorithms for interaction and affinity prediction suffer from a sharp decline in performance when handling ligands bound with novel unseen proteins. *Methods:* We propose IPBind, a geometric deep learning-based computational method, enabling robust predictions by leveraging interatomic potential between complex's bound and unbound status. *Results:* Experimental results on widely used binding affinity prediction benchmarks demonstrate the effectiveness and universality of IPBind. Meanwhile, it provids atom-level insights into prediction. *Conclusions:* This work highlight the advantage of leveraging machine learning interatomic potential for predicting protein-ligand binding affinity.**

***Index Terms*—Deep learning, drug discovery, physics-informed neural networks, protein-ligand binding affinity prediction.**

***Impact Statement*—This study extends state-of-the-art deep learning algorithms to applications in protein-ligand binding affinity prediction. This study has implications for enhancing the generalization capability of protein-ligand interactions prediction methods by interatomic potential modeling.**

I. Introduction

MODERN drug discovery is a protracted and costly endeavor, often spanning over a decade with expenditures exceeding billions of US dollars [1]. A critical bottleneck lies in identifying high-affinity ligands that bind tightly to target proteins, as binding affinity—quantified by the energy difference between unbound and bound states—serves as a pivotal indicator of therapeutic potential. Higher binding affinity reflects stronger stabilization of the bound complex relative to the separated components [2], [3]. While experimental techniques like isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR) remain gold standards, their high cost and less efficiency hinder large-scale precessing [4].

The advent of structure-based computational approaches has reshaped early-stage drug discovery by enabling rapid insilico prediction. These methods leverage 3D structural insights to rationally design ligands that complement target binding sites [5], [6]. Among them, deep learning (DL) has emerged as a transformative tool, and achieved breakthroughs in predicting protein-ligand affinities [7], [8], [9], [10]. Unlike conventional methods requiring handcrafted feature engineering, DL models enables autonomously extracting hierarchical representations from structural data and capturing intricate patterns. For example, protein-ligand complex can be represented as voxels, and employs 3D convolutional neural networks (CNN) to learn the relationship between spatial interactions and physicochemical properties [11], [12]. Modeling with graph neural networks (GNN) is a new research direction. A graph is a natural way to represent the structure of a molecule, in which each atom or residue can be set as a node and interactions between atoms can be treated as edges [13], [14], [15], [16].

Despite progress, critical challenges persist. First, the limited availability of high-quality experimental affinity data demands models with high data efficiency, which most existing DL-based

Received 11 April 2025; revised 25 August 2025 and 6 November 2025; accepted 18 January 2026. Date of publication 23 February 2026; date of current version 19 March 2026. The work of Guang Yang was supported in part by the European Research Council Innovative Medicines Initiative under Grant 101005122, in part by the Horizon 2020 program under Grant 952172, in part by the Medical Research Council under Grant MC/PC/21013, in part by the Royal Society under Grant IEC/NSFC/211235, in part by the NVIDIA Academic Hardware Grant Program, in part by the SABER project by Boehringer Ingelheim Ltd, NIHR Imperial Biomedical Research Centre under Grant RDA01, in part by the Wellcome Leap Dynamic resilience program (co-funded by Temasek Trust), in part by the UK Research and Innovation guarantee funding for Horizon Europe Marie Skłodowska-Curie Actions Postdoctoral Fellowships under Grant EP/Z002206/1, in part by a UK Research and Innovation Medical Research Council research grant, in part by TFS Research under Grant MR/U506710/1, in part by the Swiss National Science Foundation under Grant 220785, and in part by the UK Research and Innovation Future Leaders Fellowship under Grant MR/V023799/1 and Grant UKRI2738. The review of this article was arranged by Editor Paolo Bonato. *(Corresponding author: Guang Yang.)*

Krinos Li and Xianglu Xiao are with the Bioengineering Department and Imperial-X, Imperial College London, W12 7SL London, U.K.

Zijun Zhong is an independent researcher based in London, W12 7LG London, U.K.

Guang Yang is with the Bioengineering Department and Imperial-X, Imperial College London, W12 7SL London, U.K., also with the National Heart and Lung Institute, Imperial College London, SW7 2AZ London, U.K., also with the Cardiovascular Research Centre, Royal Brompton Hospital, SW3 6NP London, U.K., and also with the School of Biomedical Engineering and Imaging Sciences, King's College London, WC2R 2LS London, U.K. (e-mail: g.yang@imperial.ac.uk).

Digital Object Identifier 10.1109/OJEMB.2026.3667030

© 2026 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

PLA approaches rarely address [17]. Second, poor cross-target generalization persists, particularly when deploying models across divergent protein families. This is reflected in model's performance drops when applied to novel receptor classes beyond training domains [18], [19], [20]. Thirdly, many models overlook geometric invariance in their input representations. The lack of translational and rotational invariance in binding affinity predictions leads to inconsistent and unstable performance. Some existing works focus only on E(3) invariance, which fails to properly account for molecular chirality since E(3) network also treat reflection as invariance [21], [22].

To address these gaps, we propose IPBind, a efficient model that predicts protein-ligand binding affinity through *Interatomic Potential* modeling. Our key contributions are: (1): IPBind enables efficient training while achieving better generalization to unseen proteins. Additionally, it offers intuitive interpretability of predictions. (2) IPBind is SE(3)-invariant, making it more aligned with real-world physics. (3): Our model demonstrates robust performance across predicted structures as input, extending predictive capabilities beyond existing methods. These characteristics highlight the potential of the proposed IPBind for real-world drug discovery pipelines.

II. Related Works

A. Geometric Deep Learning in Molecular Modeling

Geometric deep learning has emerged as a powerful paradigm for modeling molecular systems, as it enables the explicit encoding of spatial relationships and symmetry properties. Early approaches introduced E(3)-invariant or equivariant architectures, such as SchNet [23], EGNN [24], and PaiNN [25], which ensure that predictions are consistent under rotations and translations. To further capture molecular chirality and directional interactions, subsequent methods extended to SE(3) models, including SE(3)-Transformers [26], TFN [27], and LEFTNet [28]. Beyond architecture-specific designs, frame averaging has been proposed as a model-agnostic strategy that enforces desired invariance or equivariance properties while allowing the use of more expressive network structures [29], [30].

B. Machine Learning Interatomic Potential

Machine learning interatomic potentials seek to model the underlying energy landscape by learning atom-wise interactions within molecular systems. In this paradigm, geometric deep learning methods naturally serve as the backbone by representing each atom and its local neighborhood [31], [32], [33]. Such models often predict per-atom contributions to the total system energy, enabling accurate estimation of binding or conformational energetics. This line of research has demonstrated broad applicability in molecular dynamics and materials science [34], [35], [36].

III. Materials and Methods

Our objective is to predict the binding affinity of protein-ligand complexes from their structures. This is a regression task where the labels represent continuous binding affinity values.

A. Overview of IPBind

We propose IPBind, a geometric deep learning model for predicting protein-ligand binding affinity. Our approach is inspired by the principle of machine learning interatomic potentials, where the total energy of a molecular system is approximated as the sum of individual atomic contributions [23], [31], [35], [37]. We derive the binding affinity by quantifying the sum of energy difference of each atom before and after binding. Fig. 1 illustrates the overall workflow of IPBind. The model takes protein-ligand complex as input. In order to make a balance between focusing on protein-ligand interactions and computational efficiency, we first cropped the protein pocket to concentrate on the most critical interaction sites, which was defined as maximum to the closest 50 residues to ligand's heavy atoms, and we preserve all heavy atoms of selected residues and ligand. The model consists of three main modules: preprocessing, massage passing, and prediction.

To address the challenge of ensuring SE(3)-invariance for 3D molecules while avoiding the computational complexity, we employ frame averaging (FA) [29], [38]. FA introduces symmetry awareness to any architecture by averaging its outputs over a carefully selected subset of group elements, as frames. In the preprocessing module, three graphs were created for encoder models based on the input structure: the bound complex, the protein, and the ligand. We then create frames for the coordinates of each input. In the Encoder module, message passing is employed to capture interatomic interactions and iteratively update atomic representations.

The prediction module leverages encoder-generated atomic embeddings to predict the energy contribution of each atom in both its unbound and bound states. To ensure SE(3)-invariance, these predictions are averaged over the output of each frame. The final binding affinity is calculated as the sum of energy differences between the bound complex and unbound components.

B. Preprocessing Module

Frame averaging: Since our goal is to predict binding affinity—-a scalar quantity—-we focus specifically on ensuring output invariance under SE(3) transformations. We adopt the Frame Averaging (FA) transformation to canonical plane to enable the output of geometric graph neural networks to be SE(3)-invariance [29]. For molecular systems with atom positions $X \in \mathbb{R}^{n\times 3}$, we define frames via principal component analysis (PCA):

1) Compute centroid and covariance matrix:

$$t = \frac{1}{n} X^\top \mathbf{1} \in \mathbb{R}^3. \tag{1}$$

$$\Sigma = (X - \mathbf{1}t^\top)^\top (X - \mathbf{1}t^\top). \tag{2}$$

2) Solve eigenvalue problem $\Sigma u = \lambda u$ to obtain eigenvectors u_1, u_2, u_3 with $\lambda_1 > \lambda_2 > \lambda_3$
3) Define 4 frames for achieving SE(3) invariant output using signed eigenbasis:

$$F_{\mathrm{SE(3)}}(X) = \left\{ ([\alpha u_1, \beta u_2, u_1 \times u_2],\, t) \;\middle|\; \alpha, \beta \in \{-1, 1\} \right\}. \tag{3}$$

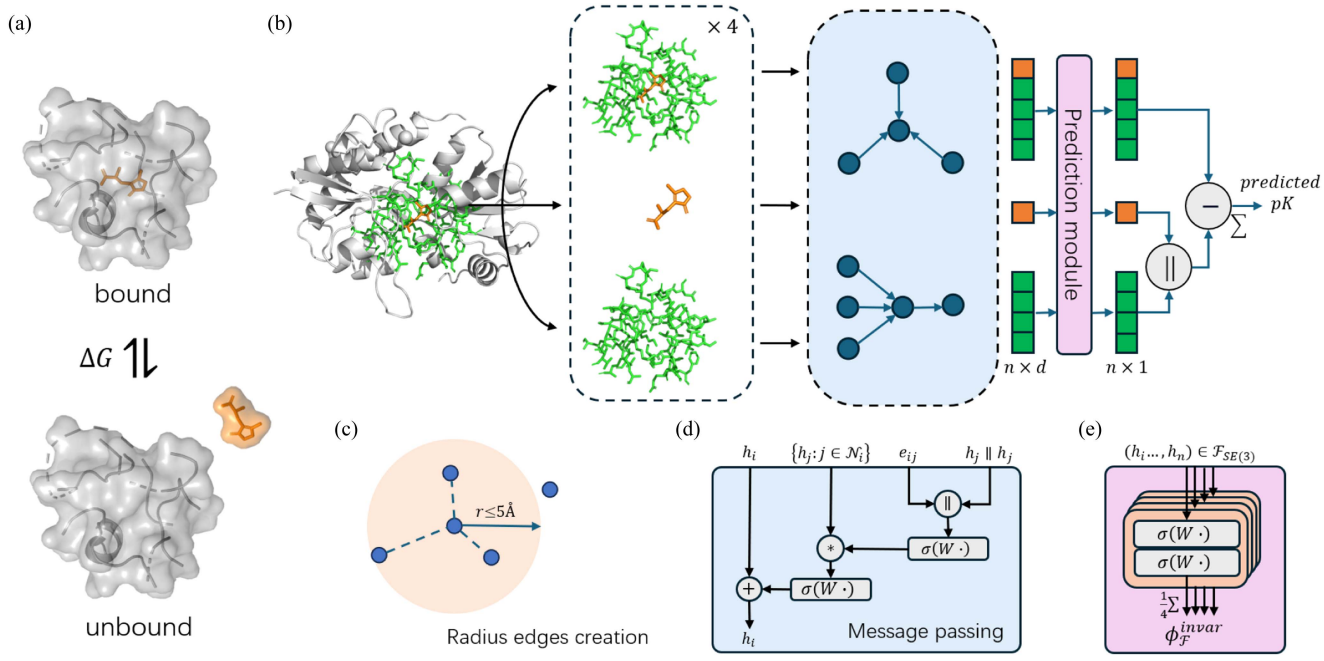


Fig. 1. Overview of IPBind (a) IPbind leverages the difference between bound and unbound status of protein-ligand to predict their binding affinity (b) Workflow of IPBind. The initial protein-ligand complex were used to construct three atom-level graphs, after encoder module, atomic contribution of each graph will be summed up to predict the binding affinity between input protein and ligand. n: number of atoms in each molecule, d: number of representation dimensions. (c) Radius edges creation (d) Message passing layer. (e) The prediction module of IPBind.

For above $(R, t) \in F_{\mathrm{SE(3)}}(X)$, we averaging the result of above transformation to achieve SE(3)-invariant by:

$$X_{\text{out}} = \frac{1}{|\mathcal{F}_{\mathrm{SE(3)}}|} \sum_{(R,t)\in\mathcal{F}_{\mathrm{SE(3)}}} \phi\left((X - \mathbf{1}t)R,\ X\right), \tag{4}$$

that is, to translation and rotation invariance of the input coordinates. This adaptation preserves the theoretical guarantees from [29] while respecting biochemical reality in chirality.

Graph creation: After frame averaging, we construct the geometric graphs based on transformed position. Three graphs incorporating bound complex, ligand and pocket atoms was built for input feature processing. We use 3D coordinates to construct radius edges at the atomic level with 5Å cutoff (Fig. 1(c).), and treat each atom as a node.

C. Encoder Module

Input embedding: The initialization of node embeddings for each atom employs a multi-dimensional feature representation approach. Specifically, the atomic number is embedded into a 128-dimensional feature vector. For edge feature representation, we utilize the radial distribution function (RBF$(\cdot)$) to transform interatomic distances d_{ij} and combine it with the relative position vectors $\tilde{\mathbf{r}}_{ij}$, ultimately embedding them into a 128-dimensional representation:

$$\mathbf{e}_{ij} = \sigma\left(\mathrm{MLP}\left(\mathrm{RBF}(d_{ij}) \parallel \tilde{\mathbf{r}}_{ij}\right)\right), \tag{5}$$

where $\parallel$ represents vector concatenation, and $\sigma(\cdot)$ is swish function.

Message passing: We adopt the interatomic interaction layer proposed in FAENet [30] to propagate atom-atom interactions defined through edges as in Fig. 1(d). During the message passing progress, only the node's representations are updated. Specifically, we learn an enhanced graph convolution filter:

$$f_{ij}^{(l)} = \sigma\left(\mathrm{MLP}\left(e_{ij} \parallel \mathbf{h}_i^{(l)} \parallel \mathbf{h}_j^{(l)}\right)\right), \tag{6}$$

where $\parallel$ represents vector concatenation. $\mathbf{h}_i^{(l)} \in \mathbb{R}^d$ denotes the feature vector of atom i at layer l. The atomic features are updated through:

$$\mathbf{h}_i^{(l+1)} = \mathbf{h}_i^{(l)} + \mathrm{MLP}\left(\sum_{j\in\mathcal{N}_i} \mathbf{h}_j^{(l)} \odot f_{ij}^{(l)}\right), \tag{7}$$

where $\odot$ denotes element-wise multiplication. $\mathcal{N}_i$ is the neighborhood of atom i.

D. Prediction Module

A two-layer multilayer perceptron is used as the output head to predict the contribution of each atom in the prediction module. After that, we average the output of 4 frames for SE(3)-invariant prediction as in Fig. 1(e). The predicted binding affinity is then calculated as the difference between the sum of atomic potentials from the unbound protein and ligand graphs and the sum from the protein-ligand complex graph.

E. Model Training and Optimization

Given the regression nature of our task, we adopt the Balanced MSE loss [39], which explicitly accounts for label distribution skewness:

$$\begin{aligned}\mathcal{L}_{\text{BME}} &= -\log p_{\text{train}}(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}) \\ &= -\log\left[\frac{p_{\text{bal}}(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})\cdot p_{\text{train}}(\boldsymbol{y})}{\int_{\mathcal{Y}} p_{\text{bal}}(\boldsymbol{y}'|\boldsymbol{x};\boldsymbol{\theta})\cdot p_{\text{train}}(\boldsymbol{y}')d\boldsymbol{y}'}\right] \\ &\cong -\log \mathcal{N}(\boldsymbol{y};\boldsymbol{y}_{\text{pred}},\sigma^2_{\text{noise}}\mathbf{I}) \\ &\quad +\log\int_{\mathcal{Y}} \mathcal{N}(\boldsymbol{y}';\boldsymbol{y}_{\text{pred}},\sigma^2_{\text{noise}}\mathbf{I})\cdot p_{\text{train}}(\boldsymbol{y}')d\boldsymbol{y}' \qquad (8)\end{aligned}$$

The noise variance σ^2_{noise} is implemented as a learnable parameter, jointly optimized with model θ.

We also incorporated an approximate NDCG loss [40] to learn the relative binding strength of different protein-ligand complexes during training. Additionally, we applied exponential amplification to enhance its effectiveness:

$$\mathcal{L}_{\text{rank}} = \exp\left(\alpha(1-\text{NDCG})\right) - 1 - \alpha(1-\text{NDCG}). \qquad (9)$$

The total loss is defined as follows:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{BME}} + \mathcal{L}_{\text{rank}} \qquad (10)$$

We used AdamW as optimizer, and use a per-batch linear warm-up and cosine annealing strategy in training. The first two epochs serve as warm-up epochs, during which the learning rate increases linearly with the batch index from 1×10^{-6} to 0.01. We set number of message passing layers as 4 for each independent run.

IV. Results

Our experiments are mainly focus on binding affinity prediction. Since the relative ranking of binding strengths is crucial in drug discovery, we primarily report Pearson correlation between labels and predicted values. We also use Root-Mean-Square-Error (RMSE) to evaluate the accuracy.

A. Dataset

The proposed model was trained and evaluated on the refined set of PDBbind2020. We employed diverse benchmarks to evaluate IPBind. CASF2016 [41] is a commonly used benchmark for assessing binding affinity prediction performance. After removing 285 test samples, we used 5050 for training and randomly selected 500 for validation across five independent runs.

As CASF2016 shares high protein similarity with training set, which may lead to over-optimistic results [42], [43], we further adopted the LBA60&LBA30 from Atom3D [44], restricting protein sequence identity to 60% and harder 30%, respectively, to better simulate real-world scenarios. A lower sequence identity indicates a lower likelihood of protein similarity between the training and test sets. The training/validation/test splits are 3563/448/452 for LBA60 and 3507/446/490 for LBA30. All experiments were repeated five times with different random seeds following prior work [10], [19], [20].

TABLE I
PERFORMANCE AT CASF16 BENCHMARK. THE BEST ONE ISBOLDED, THE SECOND BEST IS UNDERLINED.RESULT WITH BOLD DONATES THE BEST PERFORMANCE.

Model	RMSE↓	Pearson↑
Pafnucy	1.450±0.047	0.769±0.019
SIGN	1.295±0.010	0.807±0.004
OnionNet	1.399±0.076	0.770±0.027
PotentialNet	1.503±0.033	0.772±0.007
GNN-DTI	1.384±0.013	0.779±0.008
IGN	1.269±0.030	0.821±0.013
OnionNet-2	**1.164**	**0.864**
SchNet	1.390±0.023	0.787±0.016
EGNN	1.289±0.021	0.816±0.011
MetalProGNet	1.309±0.043	0.802±0.013
GraphDTA	1.568±0.036	0.698±0.016
GIGN	1.190±0.017	0.840±0.007
SS-GNN	1.165±0.011	0.846±0.004
Tankbind	1.402±0.021	0.773±0.008
EHIGN	<u>1.150±0.022</u>	<u>0.854±0.004</u>
PSICHIC	1.336±0.031	0.792±0.011
IPBind	1.173±0.016	0.849±0.042

B. Results on CASF16

We compared IPBind with wide arrange of previous studies [11], [16], [22], [43], [45], [46] at CASF16 in Table I. The results show that IPBind performs excellently in RMSE and Pearson correlation coefficient. Specifically, IPBind outperforms other GNN-based methods such as GIGN and SS-GNN while maintaining comparable performance to the current state-of-the-art EHIGN. Notably, the moderate standard deviation in Pearson performance indicates consistent performance across diverse protein-ligand complexes.

C. Results on Dissimilar Protein Scenarios With Atom3D

The advantage of IPBind becomes more pronounced when evaluated on the challenging Atom3D LBA splits designed for rigorous generalization testing. Table II shows the results of the LBA30 and LBA60 benchmark tests of Atom3D. For protein sequence similarity of train-test are reduced to 60% identity, our model achieves an unprecedented Pearson correlation of 0.843, surpassing all baselines by $\geq$8.9%. For the hardest low-sequence-similarity split at 30% identity threshold, IPBind establishes new state-of-the-art performance with a Pearson correlation of 0.732, a 19.6% relative improvement over previous best methods. The RMSE metrics confirm these advancements, corresponding to 4.1% and 12.2% error reductions from LBA60 to LBA30 versus EHIGN respectively. Meanwhile, compare with GIGN which build for E(3) invariance, our SE(3) method achieve 25% better in Pearson and 11.9% better in RMSE. These results further indicate the importance of considering physical rationality into modeling.

We further provide a comparison of the results from above three benchmarks in fig. 2, with our IPBind and five baselines tested in all of these three benchmarks: PSICHIC, Tankbind, OnionNet-2, GIGN and EHIGN. We observed that as the test

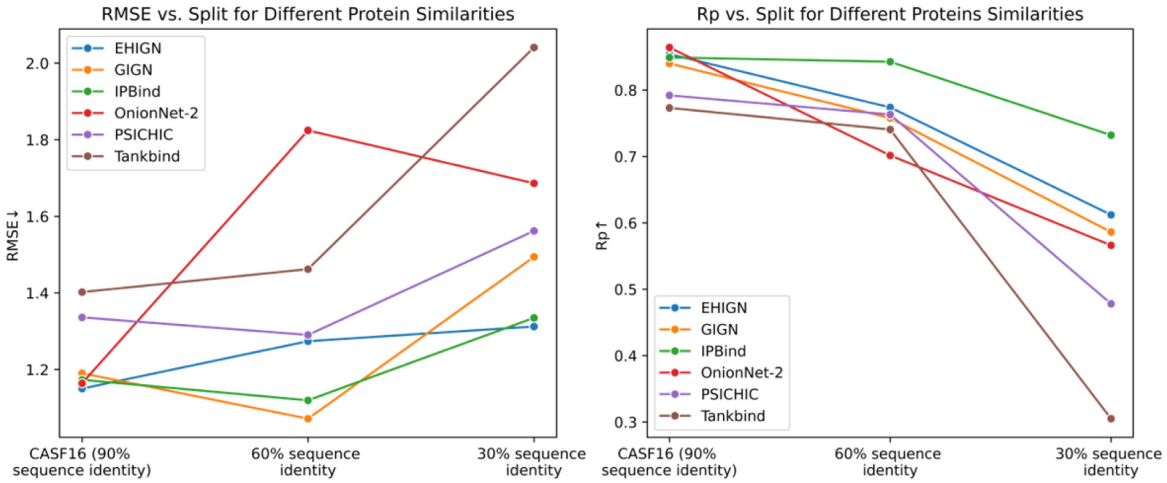


Fig. 2. Performance comparison between model performance under different level of test protein sequence identity.

TABLE II
PERFORMANCE COMPARISON AT LBA30 AND LBA60. THE BEST ONE IS BOLDED, THE SECOND BEST IS UNDERLINED. WE RAN ONIONNET-2, PSICHIC, TANKBIND, GIGN AND EHIGN USING THEIR REPORTED SETTINGS, AND THE REST OF BASELINE RESULTS ARE DIRECT CITED FROM [20].

Model	Sequence Identity 30%		Sequence Identity 60%	
	RMSE↓	Pearson↑	RMSE↓	Pearson↑
DeepDTA	1.866±0.080	0.472±0.022	1.762±0.261	0.666±0.012
B&B	1.985±0.006	0.165±0.006	1.891±0.004	0.249±0.006
TAPE	1.890±0.035	0.338±0.044	1.633±0.016	0.568±0.033
ProtTrans	1.544±0.015	0.438±0.053	1.641±0.016	0.595±0.014
MaSIF	1.484±0.018	0.467±0.020	1.426±0.017	0.709±0.008
IEConv	1.554±0.016	0.414±0.053	1.473±0.024	0.667±0.011
Holoprot	1.464±0.006	0.509±0.002	1.365±0.038	0.749±0.014
ProtNet	1.463±0.001	0.551±0.005	1.343±0.025	0.765±0.009
EGNN-PLM	1.403±0.010	0.565±0.020	1.559±0.020	0.644±0.020
Uni-Mol	1.520±0.030	0.558±0.000	1.619±0.040	0.645±0.020
ProFSA	1.377±0.010	0.628±0.010	1.377±0.010	0.764±0.000
OnionNet-2	1.695±0.004	0.534±0.000	1.802±0.022	0.705±0.001
PSICHIC	1.561±0.011	0.478±0.001	1.290±0.004	0.763±0.001
Tankbind	2.041±0.049	0.305±0.003	1.462±0.007	0.741±0.000
GIGN	1.494±0.004	0.586±0.001	**1.071±0.002**	0.757±0.001
EHIGN	**1.312±0.001**	0.612±0.000	1.274±0.002	0.774±0.000
IPBind	1.335±0.004	**0.732±0.000**	1.119±0.000	**0.843±0.000**

protein sequence identity decreases from 90% (CASF16) to 30%, the performance trends of various models become evident. This trend is consistent with the findings of previous work, suggesting that generalizing to proteins dissimilar to the training set remains a major challenge for binding affinity prediction models [18], [20]. In terms of RMSE, IPBind demonstrates a relatively stable and moderate increasing pattern. Regarding the Pearson correlation metric, IPBind maintains a relatively high value across different sequence identities. This graphical evidence, in line with the tabular results from the previous benchmarks, further validates that that the IPBind model can still maintain advanced prediction performance in difficult scenarios.

D. IPBind is Robust to Input Structures

Due to considerations of efficiency and cost, using crystal structures as model inputs is often impractical in real-world drug discovery, and predicted structures are typically employed instead. To further test the robustness of different models under practical conditions, we conducted additional tests on the more challenging LBA60 and LBA30 splits, considering two scenarios of predicted structures as inputs. The first is the lock-and-key case, where protein structure is considered a rigid body before and after binding [47]. We redocked the protein's crystal structure with ligand molecules in this case [48]. An other is the induced-fit scenario, the protein is regarded as flexible before and after binding with ligand and may differ from the known crystal structure. We illuminated the difference between these structures in fig. 3(a). To simulate this situation, we using co-folding tools to generate structures from protein sequence and ligand SMILES [49], [50].

We compared the performance of IPBind with that of PSICHIC, Tankbind, OnionNet-2, GIGN, and EHIGN, as depicted in Figs. 2(b)–(e). Tankbind utilizes separate protein and ligand inputs and therefore remains unaffected by redocking structures. Additionally, PSICHIC is a SOTA sequence-based model, which implies that it is immune to variations in the input model structures. Consequently, it serves as a baseline for comparing the robustness of other models that rely on structural inputs, and we represent it with a red dashed line in the figures.

Our findings indicate that GIGN/EHIGN showed minor declines under redocking structures, underperforming PSICHIC at 60% sequence identity but outperforming slightly at 30%. However, when using co-folded protein structures, these methods and Tankbind suffered sharp performance drops, lagging far behind PSICHIC.

In contrast, IPBind and OnionNet-2 maintained robustness, with IPBind uniquely matching or exceeding PSICHIC across conditions. Notably, IPBind led in Pearson correlation on both benchmarks, demonstrating consistent performance with imperfect input structures—a critical advantage for practical applications where crystal structures are rarely available. This robustness highlights IPBind's reliability in real-world drug discovery pipelines reliant on predicted bound structures.

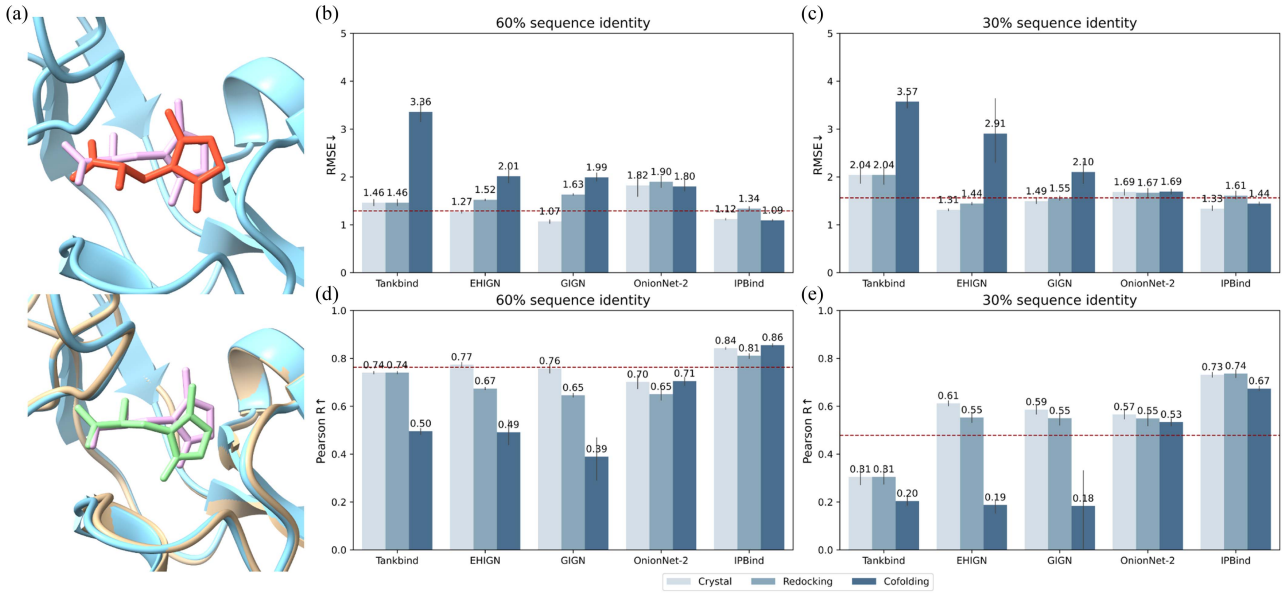


Fig. 3. Model performance under different structure input. (a) Difference between crystal structure vs redocking and cofolding structures in particular binding state with conformation PDB ID: 3DP4 (cyan: crystal protein, pink: crystal ligand, red: redocking ligand, beige: cofolding protein, green: cofolding ligand). (b) & (c) Model performance by different structure input in RMSE for LBA60 and 30. (d) & (e) Model performance by different structure input in Pearson R for LBA60 and 30. Red horizontal line indicated the performance of PSICHIC, as 1.561, 1.290, 0.478, 0.763 in (b), (c), (d), (e), respectively.

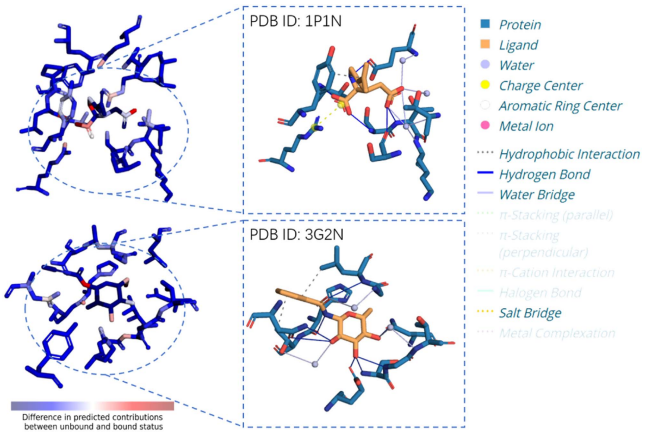


Fig. 4. IPBind predictions vs PLIP interaction visualization.In the left-side, blue indicates a lower contribution of the atom to binding affinity, while red represents a higher contribution.

E. Output Interpretability

To interpret model predictions, we visualized atomic contributions from unbound/bound difference and compared them with Protein-Ligand Interaction Profiler (PLIP [51]) analysis (Fig. 4). For a well-predicted complex (1P1N, predicted 6.46 vs. true 6.8), IPBind correctly identified key binding-region atoms consistent with PLIP. In an underpredicted case (3G2N, predicted 5.41 vs. true 4.09), the model missed water-bridge-forming atoms, highlighting specific interaction types as potential improvement areas. These results further indicate that our framework design is simultaneously provide a valuable tool for understanding model predictions. When the predicted values deviate from the true values, by comparing the atom-contribution visualizations of IPBind and PLIP, researchers can easily spot which types of interactions or which specific atoms the model has overlooked or misjudged.

V. Discussion

IPBind's design balances structural modeling and physical symmetry to improve binding affinity prediction. Ablation studies (Table III) confirm the critical role of SE(3) invariance: using E(3) or no symmetry group reduced Pearson R by up to 7.6% at 30% sequence identity, underscoring the importance of chirality-aware modeling. Modeling the difference between unbound and bound states of protein-ligand, outperforms vanilla approaches that only consider bound complex as input by 70% in Pearson R. Using shared encoders (vs. IPBind$_{3Networks}$ using separate networks for protein, ligand, and complex) also improving performance significantly. Meanwhile, we found replace protein into residue level graph IPBind$_{residue}$ could also harm the performance. Together these results validating the effectiveness of our unified architecture that leveraging molecular interactomic potentials for binding affinity prediction.

IPBind also offers superior training efficiency (Fig. 5), achieving state-of-the-art Pearson R on LBA60 within 14 minutes, and 8.9% better than competitors with longer training times. Its minimal input requirements (atomic numbers and coordinates) eliminate costly preprocessing steps like charge calculation, making it highly deployable.

TABLE III
RESULTS OF ABLATION STUDIES

Model	Group Symmetry	Proposed Framework	Sequence Identity 30%		Sequence Identity 60%	
			RMSE↓	Pearson↑	RMSE↓	Pearson↑
IPBind	SE(3)	Yes	1.335±0.066	0.732±0.013	1.119±0.021	0.843±0.006
IPBind	E(3)	Yes	1.433±0.149	0.697±0.030	1.152±0.026	0.829±0.005
IPBind	N/A	Yes	1.607±0.087	0.676±0.009	1.266±0.027	0.802±0.008
IPBind$_{residue}$	SE(3)	Yes	1.546±0.064	0.652±0.041	1.240±0.073	0.772±0.031
IPBind$_{residue}$	SE(3)	No	1.625±0.067	0.624±0.049	1.420±0.087	0.694±0.055
IPBind$_{3Networks}$	SE(3)	Yes	1.776±0.037	0.411±0.007	1.832±0.016	0.521±0.009
IPBind	SE(3)	No	2.116±0.017	0.302±0.003	1.896±0.011	0.419±0.008

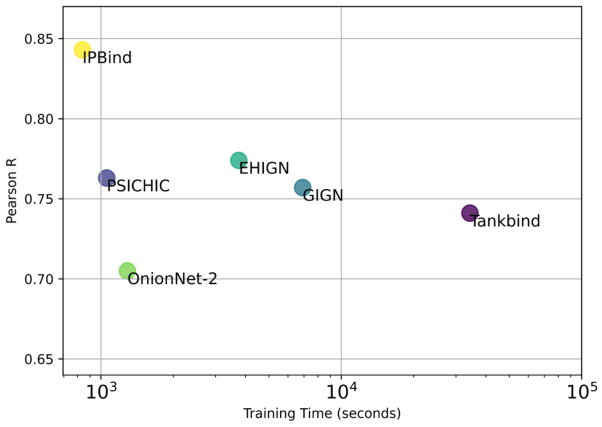


Fig. 5. Performance vs average training time over 5 runs on LBA-60 benchmark.

VI. CONCLUSION

We proposed IPBind, a SE(3)-invariant method for protein-ligand binding affinity prediction. IPBind combines high prediction accuracy, structural robustness, interpretability, and computational efficiency, providing a promising tool for protein-ligand binding affinity prediction. Its performance across diverse benchmarks and realistic scenarios positions it as a valuable advancement for applications in drug discovery.

Although IPBind demonstrates robustness to variations in binding poses and protein conformations, its performance may still be influenced by the quality of input structural graphs. Future work could also explore its generalization to increasingly diverse ligand chemotypes, which would further broaden its applicability. Future extensions of IPBind could also explore graph transformer architectures to better capture more expressive representations [52], [53]. Another promising avenue is aligning IPBind's outputs with sequence-based models, combining the generalization strength of structure-based approaches with the inference efficiency of sequence-based models.

REFERENCES

[1] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen, "Innovation in the pharmaceutical industry: New estimates of R&D costs," *J. health Econ.*, vol. 47, pp. 20–33, 2016.

[2] J. P. Hughes, S. Rees, S. B. Kalindjian, and K. L. Philpott, "Principles of early drug discovery," *Brit. J. Pharmacol.*, vol. 162, no. 6, pp. 1239–1249, 2011.

[3] I. A. Guedes, F. S. Pereira, and L. E. Dardenne, "Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges," *Front. Pharmacol.*, vol. 9, 2018, Art. no. 1089.

[4] X. Du et al., "Insights into protein–ligand interactions: Mechanisms, models, and methods," *Int. J. Mol. Sci.*, vol. 17, no. 2, 2016, Art. no. 144.

[5] P. G. Francoeur et al., "Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design," *J. Chem. Inf. Model.*, vol. 60, no. 9, pp. 4200–4215, 2020.

[6] V. T. Sabe et al., "Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review," *Eur. J. Med. Chem.*, vol. 224, 2021, Art. no. 113705.

[7] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: Deep drug–target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, 2018.

[8] T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, and S. Venkatesh, "GraphTA: Predicting drug–target binding affinity with graph neural networks," *Bioinformatics*, vol. 37, no. 8, pp. 1140–1147, 2021.

[9] K. Huang, T. Fu, L. M. Glass, M. Zitnik, C. Xiao, and J. Sun, "DeepPurpose: A deep learning library for drug–target interaction prediction," *Bioinformatics*, vol. 36, no. 22–23, pp. 5545–5547, 2020.

[10] V. R. Somnath, C. Bunne, and A. Krause, "Multi-scale representation learning on proteins," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 25244–25255.

[11] M. M. Stepniewska-Dziubinska, P. Zielenkiewicz, and P. Siedlecki, "Development and evaluation of a deep learning model for protein–ligand binding affinity prediction," *Bioinformatics*, vol. 34, no. 21, pp. 3666–3674, 2018.

[12] J. Jiménez, M. Skalic, G. Martinez-Rosell, and G. D. Fabritiis, "K Deep: Protein–ligand absolute binding affinity prediction via 3D-convolutional neural networks," *J. Chem. Inf. Model.*, vol. 58, no. 2, pp. 287–296, 2018.

[13] S. Zhang et al., "SS-GNN: A simple-structured graph neural network for affinity prediction," *ACS Omega*, vol. 8, no. 25, pp. 22496–22507, 2023.

[14] W. Lu, Q. Wu, J. Zhang, J. Rao, C. Li, and S. Zheng, "TankBind: Trigonometry-aware neural networks for drug-protein binding structure prediction," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 7236–7249.

[15] S. Lu, Z. Gao, D. He, L. Zhang, and G. Ke, "Highly accurate quantum chemical property prediction with Uni-Mol," 2023, *arXiv:2303.16982*.

[16] H. Y. Koh, A. T. Nguyen, S. Pan, L. T. May, and G. I. Webb, "Physicochemical graph neural network for learning protein–ligand interaction fingerprints from sequence data," *Nature Mach. Intell.*, vol. 6, no. 6, pp. 673–687, 2024.

[17] G. Durant, F. Boyles, K. Birchall, and C. M. Deane, "The future of machine learning for small-molecule drug discovery will be driven by data," *Nature Comput. Sci.*, vol. 4, pp. 735–743, 2024.

[18] K. Zhang et al., "Artificial intelligence in drug development," *Nature Med.*, vol. 31, pp. 45–59, 2025.

[19] L. Wang, H. Liu, Y. Liu, J. Kurtin, and S. Ji, "Learning hierarchical protein representations via complete 3D graph networks," 2022, *arXiv:2207.12600*.

[20] B. Gao et al., "ProfSA: Self-supervised pocket pretraining via protein fragment-surroundings alignment," 2023, *arXiv:2310.07229*.

[21] K. E. Wu et al., "Protein structure generation via folding diffusion," *Nature Commun.*, vol. 15, no. 1, 2024, Art. no. 1059.
[22] Z. Yang, W. Zhong, Q. Lv, T. Dong, and C. Y.-C. Chen, "Geometric interaction graph neural network for predicting protein–ligand binding affinities from 3D structures (GIGN)," *J. Phys. Chem. Lett.*, vol. 14, no. 8, pp. 2020–2033, 2023.
[23] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SCHNet–A deep learning architecture for molecules and materials," *J. Chem. Phys.*, vol. 148, no. 24, 2018, Art. no. 241722.
[24] V. G. Satorras, E. Hoogeboom, and M. Welling, "E (N) Equivariant graph neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 9323–9332.
[25] K. Schütt, O. Unke, and M. Gastegger, "Equivariant message passing for the prediction of tensorial properties and molecular spectra," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 9377–9388.
[26] F. Fuchs, D. Worrall, V. Fischer, and M. Welling, "SE (3)-Transformers: 3D roto-translation equivariant attention networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 1970–1981.
[27] N. Thomas et al., "Tensor field networks: Rotation-and translation-equivariant neural networks for 3D point clouds," 2018, *arXiv:1802.08219*.
[28] Y. Du et al., "A new perspective on building efficient and expressive 3D equivariant graph neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, vol. 36, pp. 66647–66674.
[29] O. Puny et al., "Frame averaging for invariant and equivariant network design," 2021, *arXiv:2110.03336*.
[30] A. A. Duval et al., "FAENet: Frame averaging equivariant GNN for materials modeling," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 9013–9033.
[31] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csányi, "MACE: Higher order equivariant message passing neural networks for fast and accurate force fields," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 11423–11436.
[32] Y.-L. Liao and T. Smidt, "Equiformer: Equivariant graph attention transformer for 3D atomistic graphs," 2022, *arXiv:2206.11990*.
[33] Y.-L. Liao, B. Wood, A. Das, and T. Smidt, "EquiFormerV2: Improved equivariant transformer for scaling to higher-degree representations," 2023, *arXiv:2306.12059*.
[34] Y. Wang et al., "VISNet: An equivariant geometry-enhanced graph neural network with vector-scalar interactive message passing for molecules," 2022, *arXiv:2210.16518*.
[35] S. Batzner et al., "E (3)-Equivariant graph neural networks for data-efficient and accurate interatomic potentials," *Nature Commun.*, vol. 13, no. 1, 2022, Art. no. 2453.
[36] A. Musaelian et al., "Learning local equivariant representations for large-scale atomistic dynamics," *Nature Commun.*, vol. 14, no. 1, 2023, Art. no. 579.
[37] N. Fedik et al., "Extending machine learning beyond interatomic potentials for predicting molecular properties," *Nature Rev. Chem.*, vol. 6, no. 9, pp. 653–672, 2022.
[38] A. Duval et al., "A Hitchhiker's guide to geometric GNNs for 3D atomic systems," 2023, *arXiv:2312.07511*.
[39] J. Ren, M. Zhang, C. Yu, and Z. Liu, "Balanced MSE for imbalanced visual regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 7926–7935.
[40] H. Valizadegan, R. Jin, R. Zhang, and J. Mao, "Learning to rank by optimizing NDCG measure," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, vol. 22.
[41] M. Su et al., "Comparative assessment of scoring functions: The CASF-2016 update," *J. Chem. Inf. Model.*, vol. 59, no. 2, pp. 895–913, 2018.
[42] G. C. Kanakala, R. Aggarwal, D. Nayar, and U. D. Priyakumar, "Latent biases in machine learning models for predicting binding affinities using popular data sets," *ACS Omega*, vol. 8, no. 2, pp. 2389–2397, 2023.
[43] Z. Yang et al., "Interaction-based inductive bias in graph neural networks: Enhancing protein-ligand binding affinity predictions from 3D structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 8191–8208, Dec. 2024.
[44] R. J. Townshend et al., "Atom3D: Tasks on molecules in three dimensions," 2020, *arXiv:2012.04035*.
[45] S. Li et al., "Structure-aware interactive graph neural networks for the prediction of protein-ligand binding affinity," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2021, pp. 975–985.
[46] Z. Wang et al., "OnionNet-2: A convolutional neural network model for predicting protein-ligand binding affinity based on residue-atom contacting shells," *Front. Chem.*, vol. 9, 2021, Art. no. 753002.
[47] T. Holyoak, "Molecular recognition: Lock-and-key, induced fit, and conformational selection," in *Encyclopedia of Biophysics*. Berlin, Germany: Springer, 2013, pp. 1584–1588.
[48] J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli, "Autodock VINA 1.2. 0: New docking methods, expanded force field, and python bindings," *J. Chem. Inf. Model.*, vol. 61, no. 8, pp. 3891–3898, 2021.
[49] C. D. team et al., "CHAI-1: Decoding the molecular interactions of life," *BioRxiv*, 2024.
[50] J. Abramson et al., "Accurate structure prediction of biomolecular interactions with AlphaFold 3," *Nature*, vol. 630, no. 8016, pp. 493–500, 2024.
[51] S. Salentin, S. Schreiber, V. J. Haupt, M. F. Adasme, and M. Schroeder, "PLIP: Fully automated protein–ligand interaction profiler," *Nucleic Acids Res.*, vol. 43, no. W1, pp. W443–W447, 2015.
[52] I. Sivgin, H. A. Bedel, S. Ozturk, and T. Çukur, "A plug-in graph neural network to boost temporal sensitivity in fMRI analysis," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 9, pp. 5323–5334, Sep. 2024.
[53] A. Shehzad et al., "Graph transformers: A survey," 2024, *arXiv:2407.09777*.