PLA-MGRA: Multi-Granularity and Relation-Aware Learning for Efficient and Generalizable Protein-Ligand Binding Affinity Prediction

Shunfan Li[1], Jiangkai Long[1], Xin Zou[1], Chang Tang[2*], Yuanyuan Liu[1], Xiao He[1], Xuesong Yan[1]

[1]School of Computer Science, China University of Geosciences, Wuhan 430074, China
[2]School of Software Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
lishunfan@cug.edu.cn, tangchang@hust.edu.cn

Abstract

Protein-Ligand Affinity (PLA) prediction quantifies the interaction strength to guide rational drug design. Existing approaches typically analyze interaction at a single granularity and overlook tightly coupled relationships between protein and ligand in both structure and functionality, consequently yielding suboptimal representations, leading to significant performance drops in real-world scenarios. To address this problem, we propose **PLA-MGRA**, a minimalist and effective PLA prediction framework. Specifically, PLA-MGRA captures both fine-grained atomic details and coarse-grained functional semantics within the 3D structure of protein–ligand complexes, through *multi-granularity learning*. To further parse the coupled protein–ligand relationships, we design *relation-aware learning* to enhance the binding nature of representations. Extensive experiments demonstrate that our method achieves state-of-the-art performance on multiple protein–ligand affinity prediction benchmarks, while also offering generalizability and interpretability.

Introduction

Protein–ligand interactions underpin essential cellular processes such as gene regulation, signal transduction, and metabolism (Du et al. 2016). Accurate prediction of the binding affinity between proteins and ligands remains a central challenge in structure-based drug design (Pawson and Nash 2003; Scott et al. 2016). Traditional wet-lab assays offer high precision, but are time consuming and costly, limiting their scalability (Bleicher et al. 2003; Inglese and Auld 2007; Mayr and Bojanic 2009). In contrast, computational chemistry approaches are grounded in rigorous physical and chemical principles but struggle with high computational costs, the vast conformational space of complexes, and limited scoring function accuracy (Lim et al. 2021; Su et al. 2018; Trott and Olson 2010). Consequently, achieving high predictive accuracy while improving computational efficiency remains a key bottleneck in protein–ligand affinity prediction. With the increasing availability of high-quality protein–ligand complex structures and corresponding affinity measurements, researchers have turned their attention to data-driven approaches for affinity prediction (Yang et al. 2019; Stepniewska-Dziubinska, Zielenkiewicz, and Siedlecki 2018; Vefghi, Rahmati, and Akbari 2025; Zou et al. 2023, 2025, 2024; Deng et al. 2025b,a). Depending on whether atomic interactions and 3D structure are utilized, existing machine learning methods can be broadly categorized into *interaction-free* and *interaction-based* methods.

Interaction-free methods (Li, Zhao, and Li 2021; Öztürk, Özgür, and Ozkirimli 2018; Nguyen et al. 2021; Chen et al. 2023; Bai et al. 2023) rely solely on modeling the individual features of proteins and ligands, without considering their interactions. Such simplicity implies an intentional omission of the fundamental origin of binding affinity. Although interaction-free methods may achieve high predictive accuracy by learning statistical correlations, correlation does not imply causation (Zhang et al. 2022; Sieg, Flachsenberg, and Rarey 2019).

In contrast, interaction-based methods (Yang et al. 2024; Jiang et al. 2021; Sánchez-Cruz et al. 2021; Stepniewska-Dziubinska, Zielenkiewicz, and Siedlecki 2018; Torng and Altman 2019; Feinberg 2018) explicitly employ the 3D structure at the atomic level of protein-ligand complexes and model their interactions, aligning more closely with the physical principles that underlie the binding affinity. Representative models include physics-informed neural networks (Moon et al. 2022), and geometric interaction-based graph neural networks (Yang et al. 2023). By incorporating features directly linked to binding mechanisms, yielding better predictive accuracy and computational efficiency.

Although interaction-based methods naturally align with the mechanisms of protein-ligand binding and have become mainstream, their generalization ability and interpretability remain limited due to two key challenges:

Neglect of cooperative functional group effects. Although interaction-based methods typically represent molecules at the atomic level, the functional cooperativity among atoms is overlooked. For example, delocalized π-electrons from lone pairs on sp^2-hybridized carbons in aromatic rings can form π–π stacking with ligands, enhancing binding stability (Zhao et al. 2015). Such cooperative effects are intentionally omitted in a single atomic granularity.

Rigid assumptions of binding mechanisms. Many models follow rigid lock-and-key paradigm (Lim et al. 2021; Xu et al. 2024), and represent proteins and ligands via two irrelevant branches. However, as Koshland's induced fit theory

*Corresponding author

suggests, protein binding sites may undergo conformational changes in response to ligand interaction, enabling dynamic complementarity (Koshland Jr 1995). Irrelevant representations overlook their tightly coupled relation- ships in both structure and functionality, consequently yielding suboptimal representations for PLA prediction.

To address these challenges, we propose an interaction-based model, named PLA-MGRA, grounded in two profound insights: (1) Although atomic-level interactions drive binding affinity, the synergistic behavior of atoms within functional groups also plays a crucial role. Thus, we introduce a *multi-granularity learning* to integrate both fine-grained atomic features and coarse-grained functional semantics. (2) Beyond separately representing proteins and ligands, it is crucial to model their interactions, particularly the induced-fit-like conformational changes that occur upon binding, as these dynamic adaptations are closely linked to binding affinity. To this end, a *relation-aware learning* is applied to model the coupled relationships between the protein and the ligand. The key contributions of PLA-MGRA are summarized as follows:

- **Multi-granularity learning:** PLA-MGRA integrates fine-grained atomic details with coarse-grained functional semantics, enabling more expressive representations of proteins and ligands individually.
- **Relation-aware learning:** By modeling interaction features, PLA-MGRA effectively captures the structural and functional coupling between proteins and ligands, producing representations that better correlate with binding affinity.
- **Superior generalizability and interpretability:** Extensive experiments on three PLA datasets demonstrate that PLA-MGRA achieves SOTA, while maintaining strong generalizability and interpretability.

Related Work

In this section, we first introduce the interaction-free methods and interaction-based methods for PLA prediction. Then, we make comparisons with some methods that are closely related to our work.

Interaction-free Methods

Interaction-free methods treat affinity prediction as a purely data-driven problem, neglecting inductive biases embedded in protein–ligand interactions. For instance, DeepDTA (Öztürk, Özgür, and Ozkirimli 2018) employs two separate convolutional neural networks (CNNs) to extract features from protein and ligand sequences independently. In contrast, GraphDTA (Nguyen et al. 2021) and MGraphDTA (Yang et al. 2022) extend this approach by representing proteins and ligands as 2D molecular graphs, thus capturing structural and topological information that sequence-based representations inherently lack. To enhance the expressiveness of protein sequence representations, certain methods incorporate prior structural or physicochemical information. DGraph (Jiang et al. 2020), for example, utilizes two-dimensional amino acid distance matrices, whereas CAPLA (Jin et al. 2023) integrates secondary structure information and residue-level chemical descriptors. Although these methods often achieve high accuracy on benchmark datasets, they typically focus on fitting data rather than modeling the underlying interaction mechanisms. As a result, their generalization ability and interpretability remain limited (Karimi et al. 2020).

Interaction-based Methods

Interaction-based methods focus on the underlying mechanisms of affinity formation by leveraging the 3D atomic structures of protein-ligand complexes. Early interaction-based methods relied on domain knowledge and handcrafted descriptors to represent molecular structures and interactions (Deng, Chuaqui, and Singh 2004; Ballester and Mitchell 2010; Sánchez-Cruz et al. 2021). More recently, deep learning-based models have emerged, primarily employing 3D convolutional neural networks (3D-CNNs) and graph neural networks (GNNs). These models not only consider atomic-level interactions but also incorporate features related to chemical bonds and spatial geometry, enabling the model to learn the topology and 3D spatial features of the complexes. Kdeep (Jiménez et al. 2018) and Pafnucy (Stepniewska-Dziubinska, Zielenkiewicz, and Siedlecki 2018) apply 3D convolutions to voxelized protein-ligand complexes, extracting spatial features from discretized atomic grids. However, voxelization may cause geometric distortions during the discretization of continuous atomic coordinates, potentially misrepresenting critical interaction sites such as hydrogen bonds, negatively impacting binding affinity prediction accuracy (Kuzminykh et al. 2018). In contrast, interaction-based GNN approaches have recently gained prominence in PLA prediction because of their conformity with biochemical structures. Methods such as PotentialNet (Moon et al. 2022), IGN (Jiang et al. 2021), and GIGN (Yang et al. 2023) typically follow a three-stage modeling workflow: modeling intramolecular covalent interactions, incorporating intermolecular noncovalent interactions, and finally learning unified graph-level representations using GNNs. Unfortunately, these methods focus on a single level of granularity and neglect the influence of protein–ligand coupling on feature representation.

Comparisons

Our proposed model, PLA-MGRA represents molecules at multi-granularitiy. Unlike existing methods, which often overlook critical structural details during coarse-graining, PLA-MGRA restores these fine-grained features within the same graph. Such details are essential for molecule representation because sharp changes in activity can be caused by subtle modifications (Janela and Bajorath 2023). Further, PLA-MGRA emphasizes the coupling relationships between proteins and ligands, through relation-aware learning. We deliberately place this relation-aware learning after the multi-granularity learning stage, enabling the model to capture the coupling between proteins and ligands across multiple granularities, thereby enhancing the binding nature of representations.

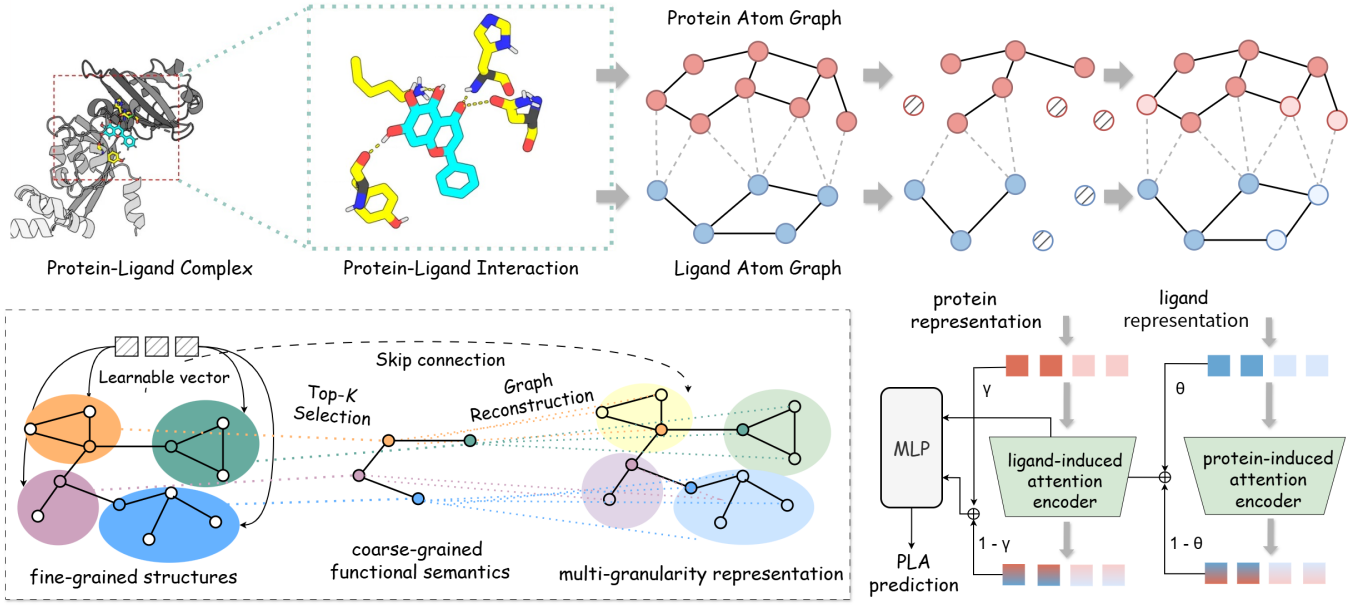

Figure 1: Architecture of PLA-MGRA for protein-ligand binding affinity prediction. To generate the input for the model, the protein-ligand complex is first modeled as a heterogeneous graph and processed using geometry-aware interaction-based graph neural network (GIGN). Then, multi-granularity learning is employed. Finally, relation-aware learning is applied to capture the coupling relationships between the protein and the ligand. We include the details of the multi-granularity learning in the bottom-left corner of the figure.

Methodology

Problem Definition. In this study, our objective is to predict the affinity of protein-ligand complexes, which is formulated as a regression problem. Each complex is represented as a graph $G = (V, E) = (V_l \cup V_p, E_l \cup E_p \cup E_{lp})$, where V_l and V_p denote the sets of nodes corresponding to atoms in the ligand and protein, respectively. Each node $v_i \in V$ is associated with a feature vector $x_i \in \mathbb{R}^d$, which encodes atomic properties, as well as a 3D coordinate $r_i \in \mathbb{R}^3$. The edge sets E_l and E_p represent covalent bonds within the ligand and protein, respectively, while the set E_{lp} captures noncovalent interactions between ligand and protein atoms. Specifically, an edge in E_{lp} is created between any pair of atoms from the ligand and the protein that are within a distance threshold of 5 Å. The target variable $y \in \mathbb{R}$ represents the binding affinity of the complex, expressed as $-\log(K_d)$, where K_d denotes the dissociation constant. The goal is to learn a predictive mapping function $f_\Theta : G \to y$, which estimates the binding affinity from the graph representation of the protein-ligand complex.

Model Architecture. The proposed PLA-MGRA architecture, as illustrated in Figure 1, consists of three components: graph encoder, multi-granularity learning, and relation-aware learning. The graph encoder is designed to extract atom-level embeddings from the complex, capturing fine-grained structural information. The subsequent multi-granularity learning operates on both the protein and the ligand, transforming their fine-grained structural features into coarse-grained functional semantic representations while restoring atomic-level details. Based on the multi-granularity features, the relation-aware learning captures the coupling relationships between the protein and the ligand, reconstructing features that are highly relevant to affinity. Ultimately, these representations are utilized to predict affinity.

Graph Encoder for Atom-Level Embeddings. It is essential to update the embeddings of protein and ligand nodes to capture atomic-level interactions. The choice of graph encoder is flexible. Candidate include GCN (Kipf and Welling 2016), SE(3)-equivariant networks (Satorras, Hoogeboom, and Welling 2021), and SE(3)-Transformer architectures (Fuchs et al. 2020). Given the geometric equivariance and the heterogeneity between covalent interactions and noncovalent interactions, we employ a geometry-aware interaction-based graph neural network (GIGN) (Yang et al. 2023) for embedding updates. This model updates node embeddings while maintaining translation and rotation invariance in 3D space, and preserves the distinction between covalent and noncovalent information propagation. The embedding update is formulated as

$$h_i = \text{GIGN}(x_i, r_i, N(v_i)), \tag{1}$$

where $h_i \in \mathbb{R}^d$ is the updated embedding of node v_i, x_i and r_i denote the atomic feature vector and 3D coordinate, respectively, $N(v_i)$ represents the set of neighbors of v_i, and d is the embedding dimension. The resulting embedding h_i can be decomposed into protein features $\mathbf{X}_{\text{pro}}$ and ligand features $\mathbf{X}_{\text{lig}}$, which serve as inputs for subsequent multi-granularity learning.

Multi-Granularity Learning. We employ a multi-granularity learning strategy on both protein and ligand atom-level embeddings to extract coarse-grained functional semantics while preserving fine-grained structural details:

$$\mathbf{X}_{\text{pro}}^{\text{MG}} = \text{MultiGran}(\mathbf{X}_{\text{pro}}, \mathbf{A}_{\text{pro}}), \tag{2}$$

$$\mathbf{X}_{\text{lig}}^{\text{MG}} = \text{MultiGran}(\mathbf{X}_{\text{lig}}, \mathbf{A}_{\text{lig}}), \tag{3}$$

where $\mathbf{X}_{\text{pro}}$ and $\mathbf{X}_{\text{lig}}$ denote the atom-level representations of the protein and ligand, respectively, and $\mathbf{A}_{\text{pro}}$ and $\mathbf{A}_{\text{lig}}$ are their corresponding adjacency matrices. Specifically, we introduce a two-stage scheme consisting of *coarse-grained abstraction* and *fine-grained restoration*. In the coarse-grained stage, the model dynamically selects the top-k most informative nodes to construct a condensed semantic graph (Gao and Ji 2019). Subsequently, during fine-grained restoration, structural details are recovered based on the coarse-grained semantics, thereby enhancing the granularity and richness of the original representations.

During coarse-grained abstraction, all nodes are ranked according to their responses to a learnable projection vector $\mathbf{p}$. The top-k nodes with the highest responses are selected to form a smaller, semantically focused subgraph. The selection process is formulated as:

$$\text{idx} = \text{TopK}\left(\frac{\mathbf{X}\cdot\mathbf{p}}{\|\mathbf{p}\|},\ k\right), \quad \tilde{\mathbf{y}} = \sigma\left(\left(\frac{\mathbf{X}\cdot\mathbf{p}}{\|\mathbf{p}\|}\right)_{\text{idx}}\right), \tag{4}$$

where $\mathbf{X}$ denotes the node feature matrix, $\tilde{\mathbf{y}}$ is a gating vector that reflects the importance of the selected nodes in the semantic space, and k is a predefined hyperparameter. Here, $\sigma(\cdot)$ denotes the sigmoid activation function. The coarse-grained representation is obtained by extracting the subgraph indexed by idx and reweighting the node features using the learned gating values:

$$\mathbf{X}^{\text{coarse}} = \mathbf{X}_{(\text{idx},:)} \odot (\tilde{\mathbf{y}} \cdot \mathbf{1}_C^\top), \quad \mathbf{A}^{\text{coarse}} = \mathbf{A}_{(\text{idx},\text{idx})}. \tag{5}$$

During fine-grained restoration, we utilize the node indices idx obtained from the coarse-grained stage to reconstruct the full-resolution graph. This process is defined as:

$$\mathbf{X}^{\text{restore}} = \text{GCN}\left(\text{distribute}\left(\mathbf{0}_{N\times C},\ \mathbf{X}^{\text{coarse}},\ \text{idx}\right)\right), \tag{6}$$

where the operation distribute$(\cdot)$ inserts the features from $\mathbf{X}^{\text{coarse}}$ into the corresponding rows of a zero-initialized matrix, guided by the indices idx, while leaving other rows unchanged. A graph convolutional network (GCN) is then applied to propagate the coarse-grained semantics throughout the full graph, thereby recovering fine-grained structural details. To combine the restored features with the original node representations to enhance both structural fidelity and semantic completeness, we incorporate a skip connection that fuses the restored representation with the original input:

$$\mathbf{X}^{\text{MG}} = \text{SkipConnect}\left(\mathbf{X}^{\text{restore}},\ \mathbf{X}\right), \tag{7}$$

where SkipConnect$(\cdot)$ denotes a residual fusion operation.

Relation-Aware Learning. Relation-aware learning is designed to obtain coupling relationships between protein and ligand. In contrast to previous methods, our approach is based on multi-granularity features, enabling both global and detailed perception of the coupling relationships. For optimization of protein representation, we first aggregated protein embeddings to get *induced protein representation* $\mathbf{X}_{\text{pro}}^{\text{induced}}$ using a *ligand-induced attention encoder*, inspired by the concept of induced fit (Koshland Jr 1995). Specifically, we use the ligand features $\mathbf{X}_{\text{lig}}^{\text{MG}}$ as the query, and the protein features $\mathbf{X}_{\text{pro}}^{\text{MG}}$ as the key and value. For brevity, we omit the superscript "MG" (multi-granularity). That is, we use $\mathbf{X}_{\text{pro}}$ to represent $\mathbf{X}_{\text{pro}}^{\text{MG}}$ and $\mathbf{X}_{\text{lig}}$ to represent $\mathbf{X}_{\text{lig}}^{\text{MG}}$ throughout the rest of this paper. The induced protein representation $\mathbf{X}_{\text{pro}}^{\text{induced}}$ is computed as:

$$\mathbf{X}_{\text{pro}}^{\text{induced}} = \text{Softmax}\left(\frac{\mathbf{X}_{\text{lig}}\mathbf{W}_Q(\mathbf{X}_{\text{pro}}\mathbf{W}_K)^\top}{\sqrt{d}}\right)(\mathbf{X}_{\text{pro}}\mathbf{W}_V). \tag{8}$$

Then, we define the final ***R**elation-**A**ware protein representation based on **M**ulti-**G**ranularity* $\text{X}_{\text{pro}}^{\text{RA-MG}}$ as:

$$\mathbf{X}_{\text{pro}}^{\text{RA-MG}} = \gamma \cdot \mathbf{X}_{\text{pro}}^{\text{MG}} + (1-\gamma)\cdot \mathbf{X}_{\text{pro}}^{\text{induced}}, \tag{9}$$

where γ is a hyperparameter used to balance the contributions of the protein features before and after induction. Since the interaction between the protein and ligand is mutual, we need to apply a symmetric operation by using a *protein-induced attention encoder*:

$$\mathbf{X}_{\text{lig}}^{\text{induced}} = \text{Softmax}\left(\frac{\mathbf{X}_{\text{pro}}\mathbf{W}'_Q(\mathbf{X}_{\text{lig}}\mathbf{W}'_K)^\top}{\sqrt{d}}\right)(\mathbf{X}_{\text{lig}}\mathbf{W}'_V), \tag{10}$$

where $\mathbf{X}_{\text{lig}}^{\text{induced}}$ is the ligand representation induced by the protein. We integrate it with the multi-granularity ligand features $\mathbf{X}_{\text{lig}}^{\text{MG}}$ to obtain the final ***R**elation-**A**ware ligand representation based on **M**ulti-**G**ranularity* $\text{X}_{\text{lig}}^{\text{RA-MG}}$:

$$\mathbf{X}_{\text{lig}}^{\text{RA-MG}} = \theta \cdot \mathbf{X}_{\text{lig}}^{\text{MG}} + (1-\theta)\cdot \mathbf{X}_{\text{lig}}^{\text{induced}}, \tag{11}$$

where θ is a hyperparameter used to balance the contributions of the ligand features before and after induction.

Affinity Prediction. After the relation-aware feature reconstruction module, we combine $\text{X}_{\text{pro}}^{\text{RA-MG}}$ and $\text{X}_{\text{lig}}^{\text{RA-MG}}$ to obtain ***R**elation-**A**ware and **M**ulti-**G**ranularity* complex feature $\text{X}_{\text{complex}}^{\text{RA-MG}}$. Specifically, we combine $\mathbf{X}_{\text{pro}}^{\text{RA-MG}}$ and $\mathbf{X}_{\text{lig}}^{\text{RA-MG}}$ using a multi-layer perceptron (MLP) and a residual connection:

$$\mathbf{X}_{\text{complex}}^{\text{RA-MG}} = \text{MLP}\left(\mathbf{X}_{\text{pro}}^{\text{RA-MG}} + \mathbf{X}_{\text{lig}}^{\text{RA-MG}}\right). \tag{12}$$

Finally, this combined representation is passed through an MLP-based regressor f_{reg} to predict the binding affinity:

$$\hat{y} = f_{\text{reg}}(\mathbf{X}_{\text{complex}}^{\text{RA-MG}}). \tag{13}$$

Model Training. To train our model, we utilize the mean squared error (MSE) loss, which measures the squared L2 norm between the predicted and actual binding affinity values. Formally, the MSE is calculated as:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2, \tag{14}$$

where $\hat{y}_i$ is the predicted binding score for the i-th protein–ligand pair, y_i is the corresponding ground truth, and n denotes the number of training examples.

Model	2013 Core Set ($N = 107$)		2016 Core Set ($N = 285$)		2019 Holdout Set ($N = 4366$)	
	RMSE ↓	R_p ↑	RMSE ↓	R_p ↑	RMSE ↓	R_p ↑
Interaction-free methods						
DeepDTA	1.639 (0.026)	0.718 (0.014)	1.357 (0.015)	0.785 (0.007)	1.485 (0.023)	0.586 (0.012)
GraphDTA	1.645 (0.085)	0.711 (0.036)	1.434 (0.064)	0.754 (0.025)	1.705 (0.075)	0.474 (0.028)
MGraphDTA	1.680 (0.093)	0.696 (0.046)	1.439 (0.047)	0.753 (0.022)	1.553 (0.028)	0.538 (0.013)
Interaction-based methods						
Pafnucy	1.517 (0.014)	0.783 (0.005)	1.450 (0.047)	0.769 (0.019)	1.438 (0.016)	0.612 (0.014)
OnionNet	1.583 (0.079)	0.741 (0.037)	1.399 (0.076)	0.770 (0.027)	1.510 (0.034)	0.573 (0.014)
PotentialNet	1.607 (0.027)	0.773 (0.010)	1.503 (0.033)	0.772 (0.007)	1.514 (0.028)	0.564 (0.014)
GNN-DTI	1.533 (0.084)	0.767 (0.040)	1.384 (0.013)	0.779 (0.008)	1.446 (0.006)	0.614 (0.007)
IGN	1.428 (0.020)	0.807 (0.001)	1.269 (0.030)	0.821 (0.013)	1.410 (0.015)	0.630 (0.008)
SchNet	1.570 (0.029)	0.754 (0.030)	1.390 (0.023)	0.787 (0.016)	1.522 (0.071)	0.560 (0.028)
EGNN	1.498 (0.025)	0.782 (0.015)	1.289 (0.021)	0.816 (0.011)	1.399 (0.013)	0.628 (0.010)
MetalProGNet	1.494 (0.027)	0.773 (0.012)	1.309 (0.043)	0.802 (0.013)	1.448 (0.021)	0.610 (0.009)
GIGN	1.380 (0.009)	0.821 (0.003)	1.190 (0.017)	0.840 (0.007)	<u>1.393 (0.007)</u>	<u>0.641 (0.006)</u>
AttentionSiteDTI	1.444 (0.037)	0.792 (0.014)	1.352 (0.022)	0.784 (0.008)	1.539 (0.015)	0.563 (0.004)
SS-GNN	<u>1.330 (0.011)</u>	<u>0.830 (0.007)</u>	<u>1.165 (0.011)</u>	<u>0.846 (0.004)</u>	1.450 (0.006)	0.633 (0.004)
GAABind	1.488	0.772	1.297	0.803	–	–
DEAttentionDTA	1.470	0.800	1.266	0.827	–	–
CL-GNN	1.345 (0.004)	0.812 (0.003)	1.200 (0.009)	0.838 (0.004)	–	–
PLA-MGRA (ours)	**1.288 (0.016)**	**0.851 (0.006)**	**1.121 (0.015)**	**0.863 (0.004)**	**1.357 (0.010)**	**0.658 (0.004)**

Table 1: Performance comparison of PLA-MGRA and baselines on the cross-dataset evaluation. All experimental results are reported in the form of *mean (standard deviation)*. The top results are shown in **bold**, and the second-best are <u>underlined</u>.

Experiments

In this section, we evaluate PLA-MGRA on the PLA prediction task. We compare our method with both the classical and the SOTA approaches. Then, we conducted ablation studies to assess the effectiveness of each component. Finally, we conducted visualization and interpretability studies to evaluate the potential of PLA-MGRA in drug discovery.

Protein-Ligand Affinity Prediction

Datasets. We evaluated the performance of PLA-MGRA on the PDBbind datasets, which offers 3D structural information of protein-ligand complexes along with experimentally measured binding affinity. To ensure a fair comparison, we adhered to the experimental settings employed in previous studies.

Evaluation. We evaluate our model using the following two evaluation paradigms:

- **Cross-dataset evaluation:** We train and validate PLA-MGRA on the *PDBbind v2016 general set* (12,904 samples), and evaluate it on three datasets: *PDBbind v2013 core set* (107 samples), *v2016 core set* (285 samples), and *v2019 holdout set* (4366 samples). It should be noted that there is no overlap among training, validation, and testing sets. This configuration is designed to assess the model's performance in different versions of the dataset.
- **Diverse protein evaluation:** As in Atom3D (Townshend et al. 2020), the *PDBbind v2019 refined set* is split according to protein sequence identity thresholds of 30% (PLA 30%) and 60% (PLA 60%), ensuring that the test proteins have lower similarity to those in the training set. This evaluation setup is designed to assess the generalizability of the model by testing it on structurally diverse proteins.

Baselines. We compare PLA-MGRA with a variety of baseline models, including interaction-free methods and interaction-based methods. In addition, for diverse protein evaluation, we also include comparisons with several pretraining-based approaches (Elnaggar et al. 2021; Liu, Guo, and Tang 2022; Wu et al. 2023). All baselines are implemented using the source codes provided by the original works and share the same training, validation, and test sets. Each experiment is repeated three times with a different random seed. We report RMSE and Pearson correlation coefficient for both settings, with the addition of Spearman correlation in the diverse protein evaluation. The mean and standard deviation of performances are then reported based on these three independent runs.

Performances on cross-dataset evaluation. To validate the effectiveness of our proposed method, we evaluate it against various models on three test sets. Table 1 presents the mean and standard deviation of the performance of each model. The results show that our model consistently achieves the best performance across all test sets. Further, it is worth noting that although SS-GNN performs comparably to our model on the 2013 core sets and 2016 core sets, it is trained on the larger *PDBbind v2019 dataset*, while our model is trained on the smaller PDBbind v2016 general dataset. This suggests that SS-GNN's performance may be overly optimistic. Moreover, the 2019 holdout set contains samples not present in the other four datasets, simulating a "temporal split" scenario-using models trained on past structural

Model	PLA 30%			PLA 60%		
	RMSE ↓	R_p ↑	R_s ↑	RMSE ↓	R_p ↑	R_s ↑
Pre-training						
ProtTrans	1.544 (0.015)	0.438 (0.058)	0.434 (0.058)	1.641 (0.016)	0.595 (0.014)	0.588 (0.009)
GeoSSL	1.451 (0.030)	0.577 (0.020)	0.572 (0.010)	–	–	–
EGNN-PLM	1.403 (0.010)	0.565 (0.020)	0.544 (0.010)	1.559 (0.020)	0.644 (0.020)	0.646 (0.020)
Interaction-free						
DeepDTA	1.866 (0.080)	0.472 (0.022)	0.471 (0.024)	1.762 (0.261)	0.666 (0.012)	0.663 (0.015)
SSA	1.985 (0.006)	0.165 (0.006)	0.152 (0.024)	1.891 (0.004)	0.249 (0.006)	0.275 (0.008)
TAPE	1.890 (0.035)	0.338 (0.044)	0.286 (0.124)	1.633 (0.016)	0.568 (0.033)	0.571 (0.021)
Interaction-based						
Atom3D-3DCNN	1.416 (0.021)	0.550 (0.021)	0.553 (0.009)	1.621 (0.025)	0.608 (0.020)	0.615 (0.028)
Atom3D-ENN	1.568 (0.012)	0.389 (0.024)	0.408 (0.021)	1.620 (0.049)	0.623 (0.015)	0.633 (0.021)
Atom3D-GNN	1.601 (0.048)	0.545 (0.052)	0.533 (0.033)	1.408 (0.069)	0.743 (0.022)	0.743 (0.027)
IEConv	1.554 (0.048)	0.414 (0.053)	0.428 (0.032)	1.473 (0.042)	0.667 (0.011)	0.675 (0.019)
MaSIF	1.484 (0.018)	0.467 (0.045)	0.455 (0.014)	1.426 (0.017)	0.709 (0.001)	0.701 (0.001)
Holoprot-Full	1.464 (0.015)	0.500 (0.026)	0.482 (0.032)	1.365 (0.038)	0.749 (0.014)	0.742 (0.011)
Holoprot-Superpixel	1.491 (0.012)	0.491 (0.014)	0.484 (0.032)	1.416 (0.022)	0.724 (0.011)	0.715 (0.006)
ProNet-AminoAcid	1.455 (0.018)	0.500 (0.013)	0.397 (0.018)	1.397 (0.018)	0.741 (0.008)	0.734 (0.009)
ProNet-Backbone	1.458 (0.003)	0.546 (0.007)	0.550 (0.008)	1.349 (0.019)	0.764 (0.006)	0.759 (0.001)
ProNet-All-Atom	1.463 (0.005)	0.551 (0.008)	0.551 (0.008)	1.343 (0.025)	0.765 (0.009)	0.761 (0.003)
PLA-MGRA (ours)	**1.387 (0.035)**	**0.614 (0.010)**	**0.616 (0.005)**	**1.282 (0.011)**	**0.781 (0.007)**	**0.777 (0.004)**

Table 2: Performance comparison of PLA-MGRA and baselines on the diverse protein evaluation.

data to predict the binding affinity of newly released structures. This setting better reflects real-world drug discovery tasks. Under this more challenging and realistic condition, SS-GNN's performance declines noticeably, whereas our model maintains its leading performance, highlighting PLA-MGRA's superior generalization ability.

Performances on diverse protein evaluation. To further assess the generalizability of our model, we performed experiments on the PDBbind 2019 refined set. Table 2 reports the result for each model in the atom3D setting. The results demonstrate that our model consistently achieves robust performance across all evaluation metrics, regardless of the data splitting strategy. This indicates that it has learned the underlying mechanisms of protein-ligand interactions rather than merely memorizing specific complexes, which highlights its advantage in predicting binding affinities for *unseen structures*. Moreover, we observe that pre-training and interaction-based models exhibit the strongest performance, while interaction-free methods generally underperform due to their inability to capture the underlying interaction mechanisms. Compared to pre-training models, which rely on massive training data and large parameter scales (e.g., EGNN-PLM with 650 million parameters), interaction-based models effectively leverage the intrinsic inductive biases of the protein-ligand binding affinity (Yang et al. 2024), achieving competitive performance with significantly fewer parameters(e.g., PLA-MGRA with 1.83 million parameters). It demonstrates our proposed model can be an efficient alternative to large-scale pre-training approaches.

Ablation Study

To validate the effectiveness of each component in PLA-MGRA, we consider the following variants of our model.

- PLA-MGRA without multi-granularity learning (w/o MG) removes the multi-granularity learning module from the full model.
- PLA-MGRA without relation-aware learning (w/o RA) removes the relation-aware learning module from the full model.
- PLA-MGRA without multi-granularity and relation-aware learning (w/o MGRA) removes both the multi-granularity learning module and the relation-aware module from the full model, leaving only the GIGN encoder.
- PLA-MGRA without GIGN (w/o GIGN) replaces it with GCN as the GNN encoder for protein-ligand representation in the full model.

As shown in Figure 2, PLA-MGRA achieves superior predictive performance by learning multi-granularity representations of proteins and ligands, as well as their relation-aware representations. The removal of any single component leads to performance degradation across all datasets. Results of the ablation study demonstrate that each component of our model is indispensable. In addition, we have the following observations: (1) PLA-MGRA significantly outperforms the variant without the multi-granularity module (w/o MG), indicating that multi-granularity learning plays a key role in protein-ligand representations learning. (2) The variant (w/o RA) suffers a significant degradation on the PDBbind v2019 test set. Notably, the PDBbind v2019 test set adopts a more realistic "time-split" evaluation protocol. Therefore, the substantial decline in performance on this set highlights

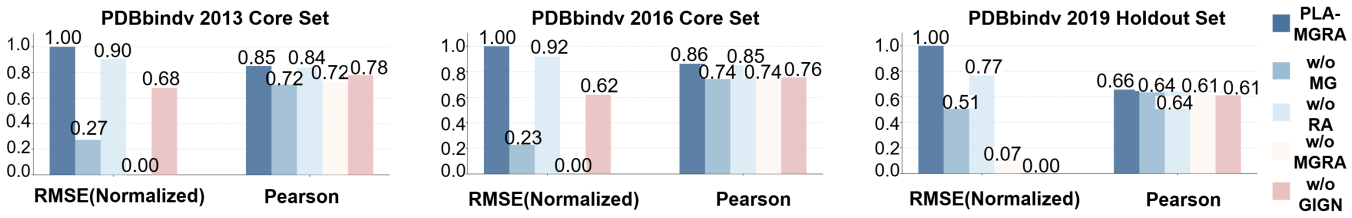

Figure 2: The performance of our model and its four variants on protein–ligand affinity prediction across three PDB datasets. To ensure consistency with the Pearson correlation coefficient (higher is better), the RMSE values are processed using inverse RMSE minmax normalization, so that higher bars in the chart indicate better performance.

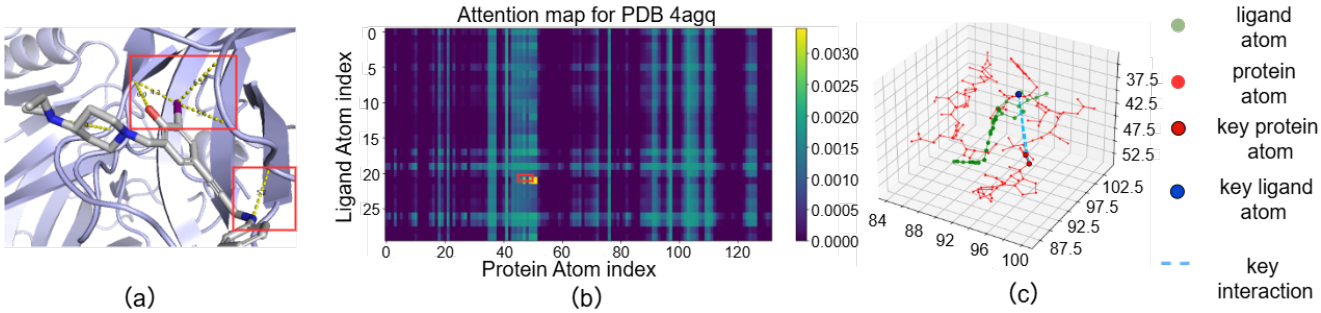

Figure 3: Visualization results of PLA-MGRA on protein-ligand complexes are presented in three parts, corresponding to (a), (b), and (c). (a) highlights key interacting atom pairs (red boxes) and their interactions (yellow dashed lines). (b) displays the atom pairs (red boxes) with critical interactions on a heatmap, illustrating the model's ability to capture critical binding regions. (c) illustrates the interactions (blue dashed lines) between key atom pairs in the 3D structure.

that relation-aware learning can enhance the model's generalization ability. (3) When both the multi-granularity learning module and relational learning module are removed (w/o MGRA), the model's performance significantly decreases and is generally worse than the variants that remove only one of these modules (w/o MG or w/o RA). This suggests that learning the relationship between proteins and ligands based on multi-granularity representations is beneficial for PLA. (4) The performance of the variant (w/o GIGN) drops noticeably compared to the full model, indicating that physical characteristics of chemical molecules, such as translation and rotation invariance, should be considered in molecular representation. These physical properties are so essential that ignoring them may hinder the model's ability, rendering subsequent processing of the original representations futile.

Visualization and Interpretability

We visualize the protein-ligand complex (PDB ID:4agq) and analyze the interpretability of the model. Figure 3 illustrates the core mechanism by which PLA-MGRA focuses on critical binding regions. It assigns higher attention weights to atom pairs known to be essential for binding, demonstrating its ability to filter out noise and identify the most relevant node features between the protein and the ligand. Further, PLA-MGRA provides intuitive 3D structural visualizations to illustrate the molecular mechanisms driving binding affinity. This interpretability makes PLA-MGRA a powerful tool for understanding protein-ligand binding processes and facilitates rational drug design.

Implementation Details

We implemented PLA-MGRA using PyTorch and PyG. All experiments were conducted on an NVIDIA TITAN V GPU with 12 GB of memory. During the feature integration stage, where original and induced features are combined, we empirically set the hyperparameters $\gamma = 0.5$, $\theta = 0.5$ based on prior validation. The model parameters were updated using the Adam optimizer with a learning rate of 1×10^{-3} and applied a weight decay of 1×10^{-6} to mitigate overfitting. The batch size was set to 128, and the model was trained for 300 epochs with early stopping.

Conclusion

In this work, we propose PLA-MGRA, an interaction-based model for PLA prediction. Benefiting from the integration of multi-granularity representations and relation-aware representations, it achieves competitive performance and demonstrates superior generalization. Moreover, PLA-MGRA exhibits high interpretability through the visualization of key atom–atom interactions, showcasing its potential in rational drug design. For example, while we have focused primarily on prediction performance, the computational efficiency and scalability of PLA-MGRA remain to be systematically evaluated. Addressing these aspects will be an important direction for our future work.

Acknowledgements

The work was supported in part by the National Natural Science Foundation of China under grants 62476258 and 62522604, and in part by the Natural Science Foundation of Hubei Province under grant 2025AFA113 and in part by the Fundamental Research Funds for National Universities, China University of Geosciences (Wuhan) under grant 2025XLB138.

References

Bai, P.; Miljković, F.; John, B.; and Lu, H. 2023. Interpretable bilinear attention network with domain adaptation improves drug–target prediction. *Nature Machine Intelligence*, 5(2): 126–136.

Ballester, P. J.; and Mitchell, J. B. 2010. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics*, 26(9): 1169–1175.

Bleicher, K. H.; Böhm, H.-J.; Müller, K.; and Alanine, A. I. 2003. Hit and lead generation: beyond high-throughput screening. *Nature reviews Drug discovery*, 2(5): 369–378.

Chen, L.; Fan, Z.; Chang, J.; Yang, R.; Hou, H.; Guo, H.; Zhang, Y.; Yang, T.; Zhou, C.; Sui, Q.; et al. 2023. Sequence-based drug design as a concept in computational drug design. *Nature communications*, 14(1): 4217.

Deng, S.; Zheng, X.; Tang, C.; Liu, X.; Liu, Y.; and An, S. 2025a. scSPAF: Cell Similarity Purified Adaptive Fusion Network for Single-Cell Multi-Omics Clustering. *IEEE Transactions on Computational Biology and Bioinformatics*.

Deng, S.; Zheng, X.; Tang, C.; Sun, K.; Liu, Y.; and Liu, X. 2025b. Find True Collaborators: Banzhaf Index-based Cross View Alignment for Partially View-aligned Clustering. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 2447–2456.

Deng, Z.; Chuaqui, C.; and Singh, J. 2004. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein- ligand binding interactions. *Journal of medicinal chemistry*, 47(2): 337–344.

Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; and Liu, S.-Q. 2016. Insights into protein–ligand interactions: mechanisms, models, and methods. *International journal of molecular sciences*, 17(2): 144.

Elnaggar, A.; Heinzinger, M.; Dallago, C.; Rehawi, G.; Wang, Y.; Jones, L.; Gibbs, T.; Feher, T.; Angerer, C.; Steinegger, M.; et al. 2021. Prottrans: Toward understanding the language of life through self-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(10): 7112–7127.

Feinberg, E. N. 2018. Debnil Sur, Zhenqin Wu, Brooke E Husic, Huanghao Mai, Yang Li, Saisai Sun, Jianyi Yang, Bharath Ramsundar, and Vijay S Pande. *Potentialnet for molecular property prediction. ACS central science*, 4(11): 1520–1530.

Fuchs, F.; Worrall, D.; Fischer, V.; and Welling, M. 2020. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in neural information processing systems*, 33: 1970–1981.

Gao, H.; and Ji, S. 2019. Graph u-nets. In *international conference on machine learning*, 2083–2092. PMLR.

Inglese, J.; and Auld, D. S. 2007. High throughput screening (HTS) techniques: applications in chemical biology. *Wiley encyclopedia of chemical biology*, 1–15.

Janela, T.; and Bajorath, J. 2023. Anatomy of potency predictions focusing on structural analogues with increasing potency differences including activity cliffs. *Journal of Chemical Information and Modeling*, 63(22): 7032–7044.

Jiang, D.; Hsieh, C.-Y.; Wu, Z.; Kang, Y.; Wang, J.; Wang, E.; Liao, B.; Shen, C.; Xu, L.; Wu, J.; et al. 2021. InteractionGraphNet: a novel and efficient deep graph representation learning framework for accurate protein–ligand interaction predictions. *Journal of medicinal chemistry*, 64(24): 18209–18232.

Jiang, M.; Li, Z.; Zhang, S.; Wang, S.; Wang, X.; Yuan, Q.; and Wei, Z. 2020. Drug–target affinity prediction using graph neural network and contact maps. *RSC advances*, 10(35): 20701–20712.

Jiménez, J.; Skalic, M.; Martinez-Rosell, G.; and De Fabritiis, G. 2018. K deep: protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. *Journal of chemical information and modeling*, 58(2): 287–296.

Jin, Z.; Wu, T.; Chen, T.; Pan, D.; Wang, X.; Xie, J.; Quan, L.; and Lyu, Q. 2023. CAPLA: improved prediction of protein–ligand binding affinity by a deep learning approach based on a cross-attention mechanism. *Bioinformatics*, 39(2): btad049.

Karimi, M.; Wu, D.; Wang, Z.; and Shen, Y. 2020. Explainable deep relational networks for predicting compound–protein affinities and contacts. *Journal of chemical information and modeling*, 61(1): 46–66.

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

Koshland Jr, D. E. 1995. The key–lock theory and the induced fit theory. *Angewandte Chemie International Edition in English*, 33(23-24): 2375–2378.

Kuzminykh, D.; Polykovskiy, D.; Kadurin, A.; Zhebrak, A.; Baskov, I.; Nikolenko, S.; Shayakhmetov, R.; and Zhavoronkov, A. 2018. 3D molecular representations based on the wave transform for convolutional neural networks. *Molecular pharmaceutics*, 15(10): 4378–4385.

Li, T.; Zhao, X.-M.; and Li, L. 2021. Co-VAE: Drug-target binding affinity prediction by co-regularized variational autoencoders. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 8861–8873.

Lim, S.; Lu, Y.; Cho, C. Y.; Sung, I.; Kim, J.; Kim, Y.; Park, S.; and Kim, S. 2021. A review on compound-protein interaction prediction methods: data, format, representation and model. *Computational and Structural Biotechnology Journal*, 19: 1541–1556.

Liu, S.; Guo, H.; and Tang, J. 2022. Molecular geometry pretraining with se (3)-invariant denoising distance matching. *arXiv preprint arXiv:2206.13602*.

Mayr, L. M.; and Bojanic, D. 2009. Novel trends in high-throughput screening. *Current opinion in pharmacology*, 9(5): 580–588.

Moon, S.; Zhung, W.; Yang, S.; Lim, J.; and Kim, W. Y. 2022. PIGNet: a physics-informed deep learning model toward generalized drug–target interaction predictions. *Chemical Science*, 13(13): 3661–3673.

Nguyen, T.; Le, H.; Quinn, T. P.; Nguyen, T.; Le, T. D.; and Venkatesh, S. 2021. GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, 37(8): 1140–1147.

Öztürk, H.; Özgür, A.; and Ozkirimli, E. 2018. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics*, 34(17): i821–i829.

Pawson, T.; and Nash, P. 2003. Assembly of cell regulatory systems through protein interaction domains. *science*, 300(5618): 445–452.

Sánchez-Cruz, N.; Medina-Franco, J. L.; Mestres, J.; and Barril, X. 2021. Extended connectivity interaction features: improving binding affinity prediction through chemical description. *Bioinformatics*, 37(10): 1376–1382.

Satorras, V. G.; Hoogeboom, E.; and Welling, M. 2021. E (n) equivariant graph neural networks. In *International conference on machine learning*, 9323–9332. PMLR.

Scott, D. E.; Bayly, A. R.; Abell, C.; and Skidmore, J. 2016. Small molecules, big targets: drug discovery faces the protein–protein interaction challenge. *Nature Reviews Drug Discovery*, 15(8): 533–550.

Sieg, J.; Flachsenberg, F.; and Rarey, M. 2019. In need of bias control: evaluating chemical data for machine learning in structure-based virtual screening. *Journal of chemical information and modeling*, 59(3): 947–961.

Stepniewska-Dziubinska, M. M.; Zielenkiewicz, P.; and Siedlecki, P. 2018. Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. *Bioinformatics*, 34(21): 3666–3674.

Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; and Wang, R. 2018. Comparative assessment of scoring functions: the CASF-2016 update. *Journal of chemical information and modeling*, 59(2): 895–913.

Torng, W.; and Altman, R. B. 2019. Graph convolutional neural networks for predicting drug-target interactions. *Journal of chemical information and modeling*, 59(10): 4131–4149.

Townshend, R. J.; Vögele, M.; Suriana, P.; Derry, A.; Powers, A.; Laloudakis, Y.; Balachandar, S.; Jing, B.; Anderson, B.; Eismann, S.; et al. 2020. Atom3d: Tasks on molecules in three dimensions. *arXiv preprint arXiv:2012.04035*.

Trott, O.; and Olson, A. J. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2): 455–461.

Vefghi, A.; Rahmati, Z.; and Akbari, M. 2025. Drug-target interaction/affinity prediction: Deep learning models and advances review. *Computers in Biology and Medicine*, 196: 110438.

Wu, F.; Tao, Y.; Radev, D.; and Xu, J. 2023. When geometric deep learning meets pretrained protein language models. *bioRxiv*, 2023–01.

Xu, X.; Xie, M.; Luo, S.; and Jia, X. 2024. Revisiting Protein–Copolymer Binding Mechanisms: Insights beyond the "Lock-and-Key" Model. *The Journal of Physical Chemistry Letters*, 15(3): 773–781.

Yang, X.; Wang, Y.; Byrne, R.; Schneider, G.; and Yang, S. 2019. Concepts of artificial intelligence for computer-assisted drug discovery. *Chemical reviews*, 119(18): 10520–10594.

Yang, Z.; Zhong, W.; Lv, Q.; Dong, T.; Chen, G.; and Chen, C. Y.-C. 2024. Interaction-based inductive bias in graph neural networks: enhancing protein-ligand binding affinity predictions from 3d structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 8191–8208.

Yang, Z.; Zhong, W.; Lv, Q.; Dong, T.; and Yu-Chian Chen, C. 2023. Geometric interaction graph neural network for predicting protein–ligand binding affinities from 3d structures (gign). *The journal of physical chemistry letters*, 14(8): 2020–2033.

Yang, Z.; Zhong, W.; Zhao, L.; and Chen, C. Y.-C. 2022. MGraphDTA: deep multiscale graph neural network for explainable drug–target binding affinity prediction. *Chemical science*, 13(3): 816–833.

Zhang, X.; Shen, C.; Liao, B.; Jiang, D.; Wang, J.; Wu, Z.; Du, H.; Wang, T.; Huo, W.; Xu, L.; et al. 2022. TocoDecoy: a new approach to design unbiased datasets for training and benchmarking machine-learning scoring functions. *Journal of Medicinal Chemistry*, 65(11): 7918–7932.

Zhao, Y.; Li, J.; Gu, H.; Wei, D.; Xu, Y.-c.; Fu, W.; and Yu, Z. 2015. Conformational preferences of π–π stacking between ligand and protein, analysis derived from crystal structure data geometric preference of π–π interaction. *Interdisciplinary Sciences: Computational Life Sciences*, 7(3): 211–220.

Zou, X.; He, X.; Zheng, X.; Zhang, W.; Chen, J.; and Tang, C. 2024. Dai-net: Dual adaptive interaction network for coordinated medication recommendation. *IEEE Journal of Biomedical and Health Informatics*.

Zou, X.; Lu, D.; Wang, Y.; Yan, Y.; Lyu, Y.; Zheng, X.; Zhang, L.; and Hu, X. 2025. Don't Just Chase" Highlighted Tokens" in MLLMs: Revisiting Visual Holistic Context Retention. *arXiv preprint arXiv:2510.02912*.

Zou, X.; Tang, C.; Zheng, X.; Li, Z.; He, X.; An, S.; and Liu, X. 2023. Dpnet: Dynamic poly-attention network for trustworthy multi-modal classification. In *Proceedings of the 31st ACM international conference on multimedia*, 3550–3559.