

# Structure-Aware Heterogeneous Information Fusion Framework for Protein–Ligand Binding Affinity Prediction

Yan Zhu, Chunyu Wang, Junjie Wang,\* Lingling Zhao,\* and Quan Zou



Cite This: *J. Chem. Inf. Model.* 2026, 66, 503–517



Read Online

ACCESS |



Metrics & More

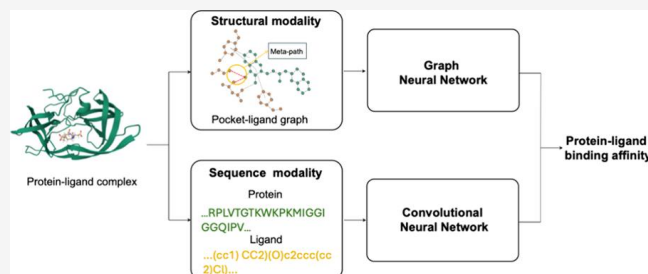


Article Recommendations



Supporting Information

**ABSTRACT:** Accurate prediction of protein–ligand binding affinities (PLAs) is essential for drug discovery and development. Recent advancements suggest that transforming protein–ligand complexes into heterogeneous graph representations may offer a viable solution. However, existing methods ignore the importance of heterogeneous graph augmentation and the complementary information provided by sequence and protein–ligand complex structure modalities, which are crucial for enhancing generalization and robustness. In this study, we propose a multimodal data fusion approach GIF-PLA (meta-path-based enhanced heterogeneous information fusion framework for protein–ligand binding affinity prediction). Protein–ligand binding complexes are represented as heterogeneous graphs with meta-paths, in parallel with protein sequences and ligand simplified molecular input line entry system (SMILES) strings, which are fed into cascaded deep neural networks, respectively. GIF-PLA effectively captures structure-oriented information, encompassing topological interactions and high-order nonlinear relationships, as well as sequence-oriented information. Finally, a late fusion module is used to integrate multilevel information. Comprehensive evaluations demonstrate that GIF-PLA surpasses state-of-the-art methods, achieving a Pearson’s correlation coefficient ( $R_p$ ) of 0.784 and a root-mean-square error (RMSE) of 1.157 on benchmark data sets. Ablation studies highlight the critical contributions of meta-paths and multimodal fusion. Overall, GIF-PLA shows significant promise in predicting protein–ligand interactions with enhanced reliability.



## 1. INTRODUCTION

Protein–ligand binding affinity (PLA) prediction is pivotal for drug discovery, bridging computational models and therapeutic development.<sup>1</sup> While experimental assays remain the gold standard, their scalability limitations have spurred computational alternatives.<sup>2,3</sup> Molecular docking tools like AutoDock Vina<sup>4</sup> and GOLD<sup>5</sup> exemplify this shift, offering rapid virtual screening through binding mode simulations, though constrained by computational cost-accuracy trade-offs.

In the pursuit of more efficient and accurate predictive models, deep learning has emerged as a promising avenue. It offers two primary methodologies for predicting protein–ligand binding affinity: complex-free and complex-based methods. Complex-free methods circumvent 3D structural dependencies by directly learning from protein sequences and ligand simplified molecular input line entry system (SMILES) representations. For example, DeepDTA<sup>6</sup> employs convolutional neural networks (CNNs) to process sequential data, whereas GraphDTA<sup>7</sup> integrates graph neural networks (GNNs) to capture ligand topological features. In contrast, complex-based methods explicitly exploit 3D structural data through divergent architectures: 3D CNNs (e.g., Pafnucy<sup>8</sup>) spatially encode atomic density grids, while GNN-based frameworks such as IGN<sup>9</sup> and EHIGN<sup>10</sup> construct heterogeneous graphs modeling geometric interactions at molecular interfaces. Notably, GNN approaches demonstrate superior predictive accuracy compared

to 3D CNNs, attributed to their inherent capacity to represent non-Euclidean molecular geometries and dynamic atomic contacts, as evidenced by EHIGN’s performance benchmarks.<sup>10</sup>

Despite the encouraging outcomes observed in previous methodologies, several obstacles and difficulties remain that impede further development. Here, we summarize three fundamental challenges that remain to be solved with regard to complex-based methods for predicting protein–ligand binding affinities:

First, despite the promising potential of transforming complexes into heterogeneous graphs for structure-based protein–ligand binding affinity prediction tasks, existing methods fail to recognize the importance of enhancing these graphs. The prevailing methodologies typically consider only first-order neighbors (i.e., directly connected nodes), thereby neglecting information from more distant nodes.<sup>9,11</sup> This limitation might impair the model’s ability to effectively capture complex structural features and semantic relationships. Meta-

**Received:** August 11, 2025

**Revised:** November 26, 2025

**Accepted:** December 2, 2025

**Published:** December 24, 2025



paths, which are a significant concept within heterogeneous graphs, may potentially address this issue; they represent ordered sequences of node types and edge types that describe composite relationships between node types.<sup>12–14</sup> Meta-paths can be viewed as higher-order proximities between two nodes, capturing semantic relationships by specifying sequences of node and edge types.<sup>15,16</sup> For example, in PLA prediction tasks, a meta-path could be “protein node–ligand node–protein node,” indicating the relationship where two protein atoms jointly interact with a ligand atom. This ability to capture higher-order proximities and semantic relationships has led to the widespread application of meta-paths in various fields.<sup>17,18</sup> However, existing heterogeneous graph-based methods often neglect to integrate the critical information extracted from meta-paths when addressing PLA tasks.<sup>9,10</sup>

Second, current research suffers from an underestimation of the importance of multimodal information fusion, particularly the integration of sequence modality and complex structure modality, in predicting PLAs. Additionally, existing multimodal methods mostly focus on the independent multimodal information on ligands and proteins separately, such as integrating protein sequences with their 3D structures, and ligand sequences with their molecular structures.<sup>19–21</sup> While leveraging multimodal information to improve PLA prediction performance has become a core research direction, and some methods have achieved notable progress in integrating structural and sequential information, how to efficiently fuse molecular modal information from different sources and at different abstraction levels remains an open question in this field.<sup>22,23</sup> This results in the fact that the potential of existing methods to fully explore and synergistically utilize the multidimensional features of proteins, ligands, and their complexes has not been fully unleashed.

Third, current PLA prediction methods exhibit significant limitations in their generalization performance, which hinders their direct application in practical virtual drug screening. Generalization performance refers to a model's ability to perform well on unseen data, which is crucial for practical applications.<sup>24,25</sup> The development of protein–ligand binding affinity prediction models is significant for tasks such as drug scoring, ranking, and screening.

To address these challenges, we developed *GIF-PLA* (meta-path-based enhanced heterogeneous information fusion framework for protein–ligand binding affinity prediction), which utilized heterogeneous graphs with meta-paths and the representations of both sequence and complex structure modalities to predict PLAs. Specifically, we first modeled the protein–ligand complexes as heterogeneous graphs, with nodes representing protein and ligand atoms and edges representing both covalent and noncovalent bonds. The graph convolutional layers were used to learn the complex interactions within the graphs. Next, two meta-paths were defined for these heterogeneous graphs: protein atom–ligand atom–protein atom and ligand atom–protein atom–ligand atom, which generated enhanced graphs for proteins and ligands. We employed graph isomorphism networks (GINs) coupled with a dynamic fusion strategy to capture the long-range dependencies and semantic relationships from these two types of graphs. In addition, protein sequences and ligand SMILES strings were encoded and analyzed with CNNs to extract fundamental composition. Finally, we integrated the multimodal information to predict PLAs, leveraging the complementary strengths of both sequence

and complex structure data. The primary contributions of this study are summarized as follows:

1. We introduce a novel graph-based heterogeneous information fusion framework for protein–ligand binding affinity prediction, which integrates multimodal data features to capture global dependencies and semantic information.
2. For affinity prediction, we are the first to integrate meta-paths into the heterogeneous graph representation of protein–ligand complexes, thereby leveraging this mechanism to enhance the model's ability to capture higher-order proximities and relevant semantic relationships.
3. We effectively combine sequence and complex structure modalities, providing a more comprehensive and robustness prediction of PLAs.

The rest of this paper is organized as follows: [Section 2](#) reviews related studies on PLA prediction methods. In [Section 3](#), we introduce the details of the network structure and its modules. [Section 4](#) details the experimental settings employed. The results and subsequent discussion are presented in [Section 5](#). Finally, conclusions are drawn in [Section 6](#).

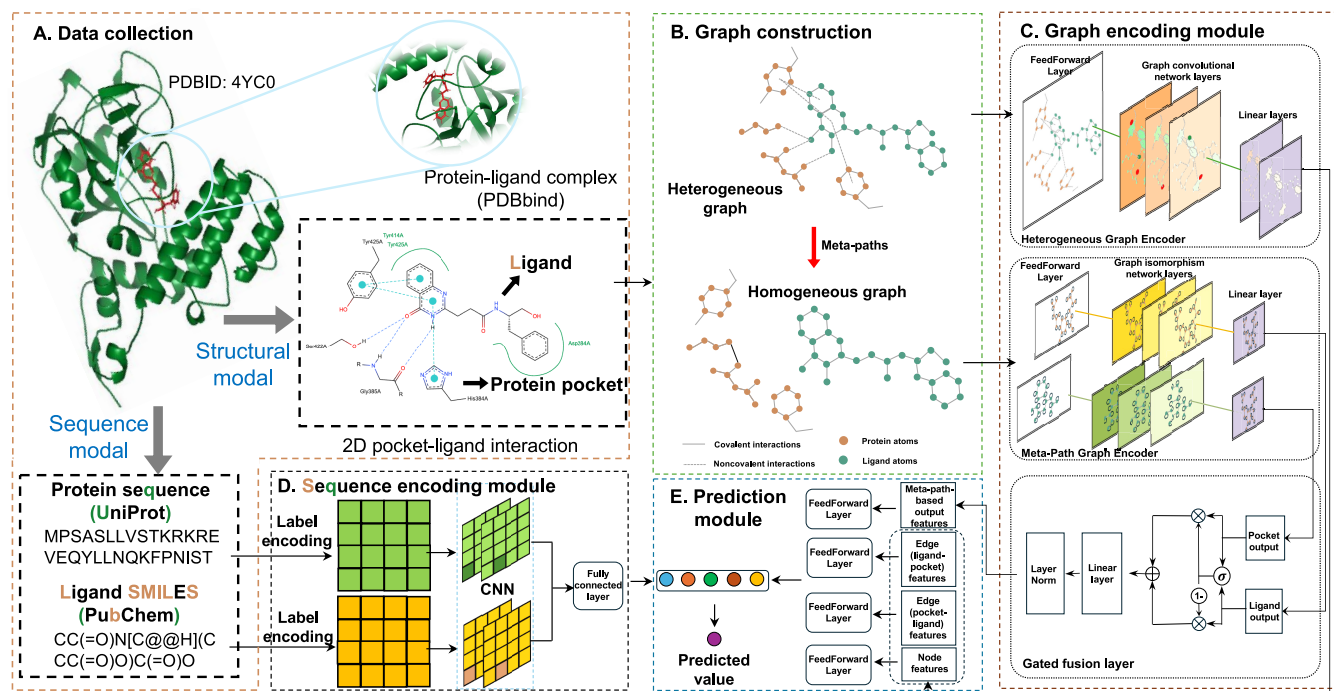
## 2. RELATED WORK

The swift progression of deep learning technologies has marked a significant leap in the prediction of PLA. The unique advantages of deep learning in processing complex data structures and feature extraction have solidified its role as an essential tool in the field of drug discovery. Current PLA prediction methods are primarily categorized into two approaches: complex-free methods and complex-based methods.

**2.1. Complex-Free Methods.** Complex-free methods do not rely on the three-dimensional structures of specific protein–ligand complexes. Instead, these methods utilize alternative types of data, such as protein sequence information, the chemical structure of ligands, and protein features (e.g., physicochemical properties, evolutionary information). The advantage of this approach lies in its ability to make effective predictions even in the absence of detailed structure information.

Many studies employed neural network frameworks to extract the features of protein sequences and drug molecules. For instance, the pioneering deep learning model DeepDTA,<sup>6</sup> which utilizes integer encoding for drug SMILES and protein amino acid sequences, followed by feature learning through a three-layer CNN. WideDTA<sup>26</sup> is an improved version of DeepDTA. To capture the latent features of drug and protein sequences for affinity prediction, Co-VAE<sup>27</sup> employs variational autoencoding technology.

Recently, transforming drug strings into molecular graphs has become a common practice. For example, GraphDTA<sup>7</sup> employs molecular graphs within GNNs, including graph convolutional networks (GCNs), graph attention networks (GATs), or GINs, to extract high-level features. For proteins, CNNs are adopted to learn hidden features. Both DeepGLSTM<sup>28</sup> and GDGRU-DTA<sup>29</sup> apply long short-term memory (LSTM) networks to learn hidden features for protein processing. For the molecular graphs, DeepGLSTM<sup>28</sup> combines multiple GCNs to learn various drug features, while GDGRU-DTA not only employs GNNs but also utilizes bidirectional gated recurrent units (GRUs) to extract drug features. TDGraphDTA<sup>30</sup> model integrates a multiscale information interaction module with a



**Figure 1.** The overall architecture of the proposed GIF-PLA. (A) Data collection: Collect data from three databases: PDBbind for structure modal data, and UniProt and PubChem for sequence modal data. (B) Graph construction: Construct heterogeneous graphs for the input pocket-ligand complex and define two meta-paths in the graph. (C) Graph encoding module: Introduce graph convolutional network (GCN) layers, graph isomorphism network (GIN) layers, and a gated fusion layer to encode the heterogeneous graphs and meta-path-based homogeneous graphs. (D) Sequence encoding module: Construct convolutional neural networks to capture the local attributes of protein sequences and ligand SMILES strings separately. (E) Prediction module: Predict binding affinities with feed-forward layers.

diffusion-driven graph optimization module that refines molecular graph representations and uncovers additional structural features. Taking drug molecular graphs and protein sequences as input, TDGraphDTA achieves accurate DTA prediction. MDCT-DTA,<sup>31</sup> based on multiscale diffusion and interaction learning, mines the node associations and structural features of drug molecular graphs through the multiscale graph diffusion convolution module, extracts the local and global features of protein sequences by means of the CNN-Transformer network module, and integrates the feature associations between drugs and proteins relying on the local interlayer information interaction structure, thereby constructing an efficient DTA prediction framework. Similarly, the Rotary Retention Graph Drug–Target Affinity network (RRGDTA<sup>32</sup>) is a DTA prediction model that integrates three core modules: MSI, which fuses multiscale drug–target information via multihead cross-attention to generate joint representations; ROE, which focuses on adjacent contexts to enhance structural features; and APM—containing IMR and SSG—which maximizes the retention of drug–target information.

Some studies utilize not only drug molecular graphs but also protein graphs to predict the binding affinities between proteins and ligands. For example, DGraphDTA<sup>33</sup> processes protein sequences to generate a contact map, which is then used to construct a graph. The protein graphs and ligand molecular graphs are then input into a GNN for feature learning. However, DGraphDTA<sup>33</sup> requires extensive database scanning due to its lengthy processing time, which significantly reduces the overall prediction accuracy. Therefore, WGNN-DTA<sup>34</sup> leverages evolutionary scale modeling (ESM) and extracts a protein graph from the model, enabling a more efficient prediction process with improved results. In addition, GEFA<sup>35</sup> transfers a

drug molecule graph to a protein graph for learning and prediction through an attention mechanism.

**2.2. Complex-Based Methods.** Complex-based methods have long been a significant approach in PLA prediction.<sup>36,37</sup> These methods utilize the 3D structure of the protein–ligand complex to estimate binding affinity based on physical interactions and scoring functions. Benchmark data sets such as CASF<sup>38</sup> and CSAR<sup>39</sup> have been crucial for evaluating the performance of these methods. The advancements in protein structure prediction through AlphaFold<sup>40,41</sup> have enhanced the accessibility of structural data, particularly for targets without experimentally resolved structures.

The 3D CNNs have become a routine usage where a 3D rectangular grid encoded with various pieces of chemical information was used as the inputs of the 3D-CNN. For instance, Pafnucy<sup>8</sup> transforms drug–target complexes into fine-grained three-dimensional grid representations, processed by 3D CNNs to generate rich feature maps, which are then integrated by fully connected layers to predict affinity values. TopologyNet<sup>42</sup> introduced topological concepts, constructing a multichannel topological neural network for predicting protein–ligand binding affinity and protein mutation stability changes. OnionNet-2<sup>43</sup> uses the logarithm of contacts between protein residues and ligand atoms as features, achieving higher predictive accuracy than previous scoring functions. HydraScreen<sup>44</sup> employs a 3D CNN in an end-to-end manner: it first samples up to 20 protein–ligand docking poses with Smina, grids the complexes, and then simultaneously outputs pose confidence and binding affinity. A differentiable PLIE ensemble score reweights the pose landscape, enabling joint optimization of pose selection and affinity prediction without requiring experimental structures. Furthermore, PointTransformer<sup>45</sup>



converts the three-dimensional structure of protein–ligand complexes into point cloud data, with each point representing an atom or group of atoms, including spatial coordinates and chemical properties, and then uses point cloud-based neural network structures, PointNet and PointTransformer, to learn data features and predict protein–ligand binding affinity.

Recently, several algorithms leveraging protein–ligand complex graphs with embedded 3D structural information have been developed, as molecular graphs effectively represent the connectivity between atoms in a molecule. For example, TopBP<sup>46</sup> introduces a topological feature extraction method that integrates multicomponent, multilevel, and electrostatic persistent homology, transforming molecular structures into topology-based images that retain chemical and biological information. These images are used by machine learning models, specifically, deep convolutional neural networks (CNNs) and ensemble tree models, for protein–ligand binding prediction and virtual screening. PLANET<sup>47</sup> builds a multiobjective graph neural network: it first extracts features from the 3D protein pocket graph and the 2D ligand graph, respectively, then lets the two representations interact via cross-attention while jointly predicting binding affinity, residue-atom contact maps, and intraligand distance matrices. ConBAP<sup>48</sup> employs EGNN, MPNN, and GVP-GNN to extract multimodal information from 3D complexes, 2D ligands, and coarse-grained protein pockets, respectively, and enhances the accuracy and generalizability of protein–ligand binding affinity prediction through a triplet contrastive learning strategy. IGN<sup>9</sup> converts the three-dimensional structure of protein–ligand complexes into graph representations, with atoms as nodes and interactions as edges, stacking two graph convolutional modules to learn intra- and intermolecular interactions within the protein–ligand complex graph. GIGN<sup>11</sup> uses 3D structures of protein–ligand complexes to generate molecular interaction representations and introduces a heterogeneous interaction layer to integrate both covalent and noncovalent interactions during the message-passing phase. Furthermore, EHIGN<sup>10</sup> processes graph-represented data through geometric interaction graph neural networks, introducing interaction-based inductive biases to better capture the complex interactions between proteins and ligands. The continuous emergence of high-quality biological structural data has ushered in new opportunities for structure-based drug-target interaction prediction research, making more precise and reliable predictions a tangible possibility.

### 3. MATERIALS AND METHODS

**3.1. GIF-PLA Architecture.** The overall flowchart of the framework is displayed in Figure 1. We first collect the data from three main databases: PDBbind<sup>49</sup> (the source of structure modal data), UniProt<sup>50</sup> and PubChem<sup>51</sup> (the source of sequence modal data) (Figure 1A). These data are divided into two branches. In the first branch, which handles structure modal data, the heterogeneous graph is constructed for the input pocket-ligand complex and two meta-paths in the graph are defined (Figure 1B). Subsequently, we introduce graph convolutional network (GCN) layers, and graph isomorphism network (GIN) layers and a gated fusion layer as the encoder module for the heterogeneous graph and meta-path-based homogeneous graphs, respectively (Figure 1C). Furthermore, a prediction module with a feed-forward layer (Figure 1D) is added to learn the representations. In the second part, which processes sequence modal data, a convolutional encoder is constructed to capture the local attributes of protein sequences

and ligand SMILES strings separately (Figure 1E). These two features are then concatenated, and fully connected layers are employed to produce the final sequence prediction output. Finally, the results from both parts are averaged to estimate the binding affinity of the protein–ligand pair. In PLA prediction, the task is to evaluate the interaction strength between a ligand and a target protein, which is typically treated as a regression problem. Given a group of data sets containing binding affinities  $(X, Y)$ , the goal is to develop a model  $F$  capable of accurately predicting the binding affinity  $\hat{y}$  for each protein–ligand pair, where  $\hat{y} \in \mathbb{R}^1$ . This can be mathematically expressed as  $\hat{y} = F(x_1, y_1)$ , where  $x_1 \in X, y_1 \in Y$ .

**3.2. Feature Extraction Module.** We calculated both sequence and structural high-dimensional features for each sample. These feature vectors will be employed as input for a deep learning model to predict binding affinity.

**3.2.1. Sequence Features.** Due to the variable lengths of protein and SMILES sequences, it is necessary to ensure fixed lengths in order to create an effective representation form. Hence, we decided on fixed maximum lengths of 1200 for protein sequences and 90 for SMILES strings based on the distribution shown in Supplementation Figure S1 so that the maximum lengths cover at least 90% of the proteins and the ligands in the data sets.<sup>6</sup> The sequences that are longer than the fixed length were truncated, whereas shorter sequences are padded with zeros.

The protein sequence  $S_p = (p_1, \dots, p_n)$ , where  $p_i$  denotes an amino acid, and the ligand SMILES strings  $S_d = (d_1, \dots, d_m)$ ,  $d_i$  denotes a ligand SMILES character, are utilized as the 1D representation inputs for another part of the model, which were encoded using label encoding. Each amino acid type is assigned a unique integer based on its associated alphabetical symbol [e.g., Alanine (A) is 1, Cystine (C) is 3, Aspartic Acid (D) is 4 and so on], allowing the protein to be represented as an integer sequence. Similarly, SMILES strings were also encoded using label encodings. For instance, the label encoding for the SMILES string “CN = C=O” would be [C N = C = O], which would correspond to [1 3 63 1 63 5] in the encoded form.

**3.2.2. Complex Structure Features.** We defined the protein pocket structure surrounding the ligand within 5 Å as a pocket-ligand complex, represented by a heterogeneous graph. This graph encompasses a protein graph and a ligand graph, as depicted in the upper half of Figure 1B (where green nodes represent the ligand atoms and orange nodes represent the protein pocket atoms). The nodes and edges in the graph are characterized by initial 2D and 3D features. Pocket nodes and ligand nodes share a set of 2D features spanning 35 dimensions, while pocket edges and ligand edges share a set of 2D features with 6 dimensions. Additionally, a set of 3D geometric features, totalling 11 dimensions and including distances, angles, and area statistics, were introduced as supplementary 3D edge features for pocket edges, ligand edges, as well as the edges from the pocket to the ligand and from the ligand to the pocket. Furthermore, two meta-path-based homographs were constructed to enhance the feature representation, as shown in the lower half of Figure 1B. We utilized two specific meta-paths: “protein atom–ligand atom–protein atom” and “ligand atom–protein atom–ligand atom” to generate homogeneous graphs for both protein and ligand. Nodes and edges retain the features from the original graph. Detailed information on all features for nodes and edges can be found in Supplementary Table S1.

**3.3. Graph Encoding Module.** **3.3.1. Heterogeneous Graph Encoder.** The heterogeneous graph encoder forms the

foundation of the GIF-PLA model, processing the protein–ligand complex as a graph  $G = (V, E)$ . Here, the node set  $V = (V_l \cup V_p)$  consists of the ligand node set  $V_l$  and the pocket node set  $V_p$ . The edge set  $E = E_l \cup E_p \cup E_{lp} \cup E_{pl}$ , where  $E_l$  and  $E_p$  are the sets of ligand and protein covalent edges (i.e., covalent bonds/interactions), respectively;  $E_{lp}$  and  $E_{pl}$  are the sets of noncovalent edges (i.e., noncovalent bonds/interactions). A noncovalent edge exists when the distance between ligand atom  $v_i \in V_l$  and protein atom  $v_k \in V_p$  satisfies the condition  $d_{ik} < \theta$ , where the distance is defined as  $d_{ik} = \|r_i - r_k\|^2$ , with a threshold set to 5 Å. This encoder begins by transforming initial node and edge features using type-specific feedforward networks:

$$h_v^{(0)} = \varphi_t(v)(x_v), \forall v \in V \quad (1)$$

$$e_{uv}^{(0)} = \psi_r(u, v)(x_{uv}), \forall (u, v) \in E \quad (2)$$

where  $\varphi_t(v)$  and  $\psi_r(u, v)$  are neural networks specific to node types and edge relations, respectively. Specifically,  $\varphi_t(v)$  is a feedforward neural network that takes the initial feature vector  $x_v$  of a node  $v$  and transforms it into an embedding space suitable for graph convolution operations. Similarly,  $\psi_r(u, v)$  is a feedforward neural network that processes the initial feature vector  $x_{uv}$  of an edge  $(u, v)$ , encoding the relationship between the connected nodes.

Following this initial encoding, the model employs  $L$  layers of heterogeneous graph convolution to update node features:

$$h_{GCN(v)}^{(l+1)} = \sigma \left( \sum_{r \in R} \sum_{u \in N_r(v)} W_r^{(l)} h_u^{(l)} + b_r^{(l)} \right), l = 0, 1, \dots, L - 1 \quad (3)$$

where  $N_r(v)$  denotes the neighbors of node  $v$  under relation  $r$ ,  $W_r^{(l)}$  and  $b_r^{(l)}$  are learnable parameters, and  $\sigma$  is a nonlinear activation function.

Notably, the model utilizes separate convolution operations for different relation types. For intramolecular interactions and intermolecular interactions, we employ a standard graph convolution operation. The key operations are defined as follows:

The message passing mechanism computes the messages sent from source nodes to destination nodes by combining the source node features and edge features. Specifically, the message  $m_v$  is computed using the ReLU activation function on the sum of the source node feature  $h_n$  and the edge feature  $h_e$ :

$$m_v = \text{ReLU}(h_n + h_e) \quad (4)$$

In the forward pass, the layer updates the node features by aggregating messages from neighboring nodes. The updated feature  $h_{new(v)}^{(l+1)}$  for node  $v$  at layer  $l+1$  is given by

$$h_{new(v)}^{(l+1)} = \text{MLP} \left( h_v^{(l)} + \sum_{u \in N(v)} m_u \right) \quad (5)$$

where  $N(v)$  is the set of neighboring nodes and  $\text{MLP}(\cdot)$  denotes a multilayer perceptron that processes the aggregated features.

**3.3.2. Meta-Path Graph Encoder.** To capture higher-order interactions mediated by the protein or ligand, GIF-PLA introduces a novel meta-path graph encoder. This encoder operates on a meta-path graph  $G_m = (V_m, E_m)$ , representing ligand atom–protein atom–ligand atom paths or protein atom–ligand atom–protein atom paths. The meta-path encoder employs a series of GIN layers:

$$h_{GIN(v)}^{(l+1)} = \text{MLP}^{(l)}((1 + \varepsilon^{(l)}) \cdot h_v^{(l)} + \sum_{u \in N(v)} h_u^{(l)}), l = 0, 1, \dots, L_m - 1 \quad (6)$$

where  $\varepsilon^{(l)}$  is a learnable parameter and  $\text{MLP}^{(l)}$  is a multilayer perceptron.

To effectively integrate the outputs from the two meta-path encoders, a gated fusion mechanism is employed. This mechanism dynamically weights the contributions of each input feature based on their relevance. The dynamic weight  $W_d$  is computed as follows:

$$w_d = \sigma(W_d \cdot (h_{\text{input1}} + h_{\text{input2}})) \quad (7)$$

In this equation,  $h_{\text{input1}}$  and  $h_{\text{input2}}$  represent the outputs from the two meta-path encoders, while  $W_d$  is a learnable weight matrix, and  $\sigma$  denotes the sigmoid activation function.

The fused output  $h_{\text{output}}$  is then calculated using the dynamic weight:

$$h_{\text{output}} = w_d \cdot h_{\text{input1}} + (1 - w_d) \cdot h_{\text{input2}} \quad (8)$$

This equation illustrates how the contributions from each input are weighted according to their importance. The final representation is further processed through a linear transformation and normalization:

$$h_{\text{final}} = \text{LayerNorm}(P \cdot h_{\text{output}}) \quad (9)$$

where  $P$  is another learnable weight matrix that facilitates the final linear transformation.

**3.4. Sequence Encoding Module.** To characterize the protein amino acid sequences  $S_p$  and drug SMILES strings  $S_d$ . These layers meticulously extract local hidden attributes, generating representation sequences  $H^{(p)}$  and  $H^{(d)}$ .

$$H^{(p)} = f_{\text{CNN}}(f_{\text{CNN}}(S_p)) \quad (10)$$

$$H^{(d)} = f_{\text{CNN}}(f_{\text{CNN}}(S_d)) \quad (11)$$

These enriched representations are then concatenated and fed into a fully connected layer, generating a prediction value that encapsulates the intricate interplay between protein sequences and drug structures:

$$y_{\text{seq}} = \sigma(w \cdot (H^{(p)}, H^{(d)}) + b) \quad (12)$$

where  $w$  and  $b$  are learnable parameters.  $\sigma$  denotes the ReLU activation function.

**3.5. Prediction Module.** The prediction module of GIF-PLA synthesizes information from both the heterogeneous graph and the meta-path graph representations to generate the final binding affinity prediction. This module incorporates edge-based predictions, node-based predictions, and meta-path predictions. Edge-based predictions aggregate intermolecular edge features and apply linear transformations:

$$h_{lp} = \text{maxPool}(\{e_{uv}: (u, v) \in E, r(u, v) = \text{inter}_{lp}\}) \quad (13)$$

$$h_{pl} = \text{maxPool}(\{e_{uv}: (u, v) \in E, r(u, v) = \text{inter}_{pl}\}) \quad (14)$$

$$y_{\text{edge1}} = \text{Linear}(h_{lp}), y_{\text{edge2}} = \text{Linear}(h_{pl}) \quad (15)$$

Node-based predictions aggregate node features separately for ligand and protein:

$$h_p = \text{maxPool}(\{h_v: v \in V, t(v) = \text{pocket}\}) \quad (16)$$

$$h_l = \text{maxPool}(\{h_v: v \in V, t(v) = \text{ligand}\}) \quad (17)$$

$$y_{\text{node}} = \text{Linear}(\text{Concat}(h_p, h_l)) \quad (18)$$

The meta-path prediction utilizes the output of the meta-path graph encoder:

$$y_{\text{meta}} = \text{Linear}(h_{G_m}) \quad (19)$$

The final prediction is an average of these diverse signals and the prediction from sequences (equal 12):

$$y_{\text{final}} = (y_{\text{edge1}} + y_{\text{edge2}} + y_{\text{node}} + y_{\text{meta}} + y_{\text{seq}})/5 \quad (20)$$

**3.6. Loss Function.** The model is trained by minimizing a mean-squared error (MSE) loss function as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (21)$$

where  $y_i$  represents the true binding affinity,  $\hat{y}_i$  represents the predicted binding affinity, and  $n$  is the number of samples.

## 4. EXPERIMENTAL SETTINGS

**4.1. Data Set Preparation.** **4.1.1. Construction of the Benchmark Data Set.** The benchmark data set is primarily derived from the PDBbind database (version 2020), encompassing the general set and refined set. This comprehensive data set includes high-quality protein–ligand complexes, each accompanied by detailed structural information and experimentally determined binding affinity data.<sup>49</sup> The biological unit of each complex is divided into a protein molecule (in PDB format) and a ligand molecule (in Mol2 and SDF formats). Binding affinities are reported using standard measurement types, including inhibition constants ( $K_i$ ), dissociation constants ( $K_d$ ), and half-maximal inhibitory concentrations ( $IC_{50}$ ).

We downloaded a total of 19,443 complexes with experimentally determined binding affinities from the general set and refined set of PDBbind (<http://www.pdbbind.org.cn/>), where the binding affinities are represented as  $-\log K_d$  or  $-\log K_i$  (with larger values indicating stronger binding). The entries were filtered based on two criteria. First, compounds should have available and valid graph representations that can be processed by RDKit, and can be successfully linked to PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) Chemical Identifiers (CIDs) with available SMILES data, resulting in the removal of 2,748 samples that could not be processed. Second, proteins can be successfully mapped to UniProt database (<https://www.uniprot.org/>) IDs with available sequence data, which led to the removal of 354 protein–ligand pairs. We utilized UniProt sequences instead of the protein sequences directly extracted from the PDB structures for the following two reasons: 1) one protein may have different sequence variations in different PDB structures; 2) the protein sequences present in the PDB structures might be incomplete, such as only including some domains or lacking some flexible regions. In total, we obtained 15,726 compound–protein pairs that satisfied these criteria. The remaining samples were randomly divided into training (11,008), validation (3,145), and test (1,573) sets in a ratio of 7:2:1 for model training, testing, and optimal model preservation. Data statistics, including overlap with the training set, are summarized in Table 1.

**4.1.2. Construction of the Additional Test Data Set.** To rigorously evaluate model generalizability, we constructed an independent external test set (Test21) by systematically

**Table 1. Detailed Data Statistics for Each Dataset**

| Data set       | Interactions | Proteins | Ligands | Protein overlap rate | Ligand overlap rate |
|----------------|--------------|----------|---------|----------------------|---------------------|
| Training set   | 11,008       | 2,340    | 8,946   | -                    | -                   |
| validation set | 3,145        | 1,164    | 2,833   | 31.95%               | 32.53%              |
| Test set       | 1,573        | 793      | 1,468   | 24.39%               | 24.82%              |
| Test21 set     | 21           | 20       | 14      | 0%                   | 0%                  |

deduplicating and filtering the CSAR-HiQ<sup>39</sup> and PL-REX<sup>52</sup> data sets against the PDBbind data. The initial combined data set contained 507 protein–ligand complexes. We retrieved amino acid sequences and SMILES strings from the UniProt and PubChem databases, respectively, based on protein and ligand names. To ensure dissimilarity from the PDBbind training data, we calculated pairwise Tanimoto similarity between all ligands in the external sets and those in the PDBbind training set using Morgan fingerprints (radius = 3, 2048 bits) in RDKit.<sup>53</sup> Ligands with a similarity score of 0.6 or higher were excluded. For proteins, we used CD-HIT-2D<sup>54</sup> with a 60% sequence identity threshold to remove proteins with significant sequence similarity to those in the PDBbind training set. These filters ensured that no compound–protein pair in the final Test21 set had greater than 60% similarity in either compound structure or protein sequence to any pair in the training set. Furthermore, structural similarity assessment via TM-score revealed minimal structural similarity between Test21 and training set proteins, with the maximum observed TM-score being only 0.425 (e.g., between Test21's 2dm5 and the training set's 4kii).

The resulting Test21 set comprises 21 unique compound–protein interactions (detailed in Table 1), with 20 unique compounds and 14 nonredundant proteins.

**4.2. Clustering-Based Data Splitting Validation.** We conducted clustering-based data splitting with the aim of assessing the model's generalizability by controlling the similarity among data sets.<sup>55,56</sup> This approach ensures that compound–protein pairs with compounds or proteins belonging to the same cluster are either all assigned to the training set, the validation set, or the test set.

We calculated the Morgan fingerprints for all ligands using the RDKit,<sup>53</sup> which were utilized to construct a similarity matrix where each element represents the similarity between corresponding ligands. The Tanimoto similarity coefficient was employed to measure the similarity among ligands. To compute the similarity between proteins, we applied the Smith–Waterman algorithm,<sup>57</sup> based on dynamic programming principles, which identifies the optimal local alignment between two sequences. By constructing the similarity matrix, we quantified the degree of similarity between each pair of ligands or proteins, as illustrated in Figure S2.

Based on the similarity matrices, the single-linkage clustering algorithm<sup>58</sup> was employed to group ligands and proteins with high similarity into clusters. A clustering threshold of 0.3 was selected for both ligands and proteins. This threshold was chosen because distances below 0.3 are insufficient for effective separation and may lead to data redundancy, while distances above 0.3 indicate lower similarity, potentially resulting in overly dispersed clusters and reducing the representativeness and information content of the entities within the clusters. Thus, a threshold of 0.3 provides a balance, ensuring similarity within clusters without causing excessive dispersion.



Data splitting was then performed based on the generated ligand and protein clusters, according to three distinct scenarios:

1. Ligand-clustering setting: Protein–ligand pairs with compounds from the same cluster cannot be shared across training, validation, and test sets. That is, for any two protein–ligand pairs from different sets, the compounds must come from different clusters.
2. Protein-clustering setting: Protein–ligand pairs with proteins from the same cluster cannot be shared across training, validation, and test sets. That is, for any two protein–ligand pairs from different sets, the proteins must come from different clusters.
3. Both-clustering setting: Both compound clusters and protein clusters cannot be shared across training, validation, and test sets. That is, for any two protein–ligand pairs from different sets, the compounds must come from different clusters, and the proteins must also come from different clusters.

All partitioning scenarios maintained a 7:2:1 ratio (training: validation: test) through cluster-wise stratified splitting, ensuring proportional representation of molecular diversity across partitions. This rigorous separation protocol creates progressively challenging evaluation scenarios, with the dual-constraint setting representing the most stringent generalization test.

**4.3. Baselines.** To comprehensively evaluate the performance of GIF-PLA and demonstrate its advantages, we compared it to a range of established baseline models using different modalities. These baselines are categorized into complex-free and complex-based methods, as described below:

1. Complex-free methods: These methods rely on protein sequence information, ligand SMILES strings, or 2D graph representations, without requiring 3D structural information on the protein–ligand complex.
2. DeepDTA:<sup>6</sup> This model employs convolutional neural networks to process protein sequences and ligand SMILES strings, offering a deep learning solution that does not require interaction data.
3. GraphDTA:<sup>7</sup> This model includes GCN, GAT, GIN, and GAT-GCN, each utilizing unique graph neural network configurations to encode molecular structures into predictive models.
4. Complex-based methods: These methods utilize the 3D structure of the protein–ligand complex to estimate binding affinity. Following the comparative methods used by EHIGN,<sup>10</sup> we included the following complex-based baselines:
5. RF-Score:<sup>37</sup> A random forest-based scoring function designed to predict protein–ligand binding affinities.
6. IGN:<sup>9</sup> A deep graph representation learning framework to learn protein–ligand interaction patterns from 3D structures, utilizing independent graph convolution modules for intramolecular and intermolecular interactions.
7. SchNet:<sup>59</sup> A deep learning architecture that allows for spatially and chemically resolved insights into quantum-mechanical observables of atomistic systems.
8. EGNN:<sup>60</sup> An E(3)-equivariant graph neural network to learn graph neural networks equivariant to rotations, translations, reflections and permutations, used for predicting molecular properties.

9. GIGN:<sup>11</sup> A geometric interaction graph neural network, which designs a heterogeneous interaction layer that unifies covalent and noncovalent interactions into the message passing phase to learn node representations more effectively. The heterogeneous interaction layer follows fundamental biological laws, including invariance to translations and rotations of the complexes, thus avoiding expensive data augmentation strategies.

10. EHIGN:<sup>10</sup> An innovative approach for protein–ligand binding affinity prediction that incorporates an interaction-based inductive bias, representing protein–ligand complexes as heterogeneous graphs and modeling binding affinity as the sum of pairwise atom–atom interactions.

**4.4. Evaluation Metrics.** To comprehensively evaluate the performance of the GIF-PLA model, we employ two primary metrics: Root Mean Square Error (RMSE) and Pearson Correlation Coefficient (Rp). These metrics provide a robust framework for assessing the accuracy and correlation of the predicted binding affinities against the true values.

RMSE is a widely used metric for measuring the average magnitude of the prediction errors. It is particularly useful in regression tasks where the goal is to minimize the difference between predicted and actual values. RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (22)$$

where  $y_i$  represents the true binding affinity,  $\hat{y}_i$  represents the predicted binding affinity, and  $n$  is the number of samples. RMSE provides a measure of how well the predicted values match the actual values, with lower RMSE indicating better model performance.

The Pearson Correlation Coefficient measures the linear correlation between the predicted and actual binding affinities. It ranges from  $-1$  to  $1$ , where  $1$  indicates a perfect positive linear relationship,  $-1$  indicates a perfect negative linear relationship, and  $0$  indicates no linear relationship. Rp is defined as

$$\text{Rp} = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (23)$$

where  $y_i$  and  $\hat{y}_i$  are the true and predicted binding affinities, respectively, and  $\bar{y}$  and  $\bar{\hat{y}}$  are their respective means. A higher PR value indicates a stronger linear relationship between the predicted and actual values, reflecting better model performance.

**4.5. Implementation of GIF-PLA.** The implementation of the GIF-PLA model is carried out using the Deep Graph Library (DGL) with PyTorch as the backend. Experiments were conducted on an NVIDIA A800 GPU with 80 GB of memory, ensuring ample computational resources for training and inference.

GIF-PLA contains 73,337,800 trainable parameters, reflecting a balanced architecture that provides sufficient capacity to capture complex molecular interactions without excessive redundancy. The computational complexity of a single forward pass is 845.386 GFLOPs, which quantifies the per-prediction computational cost. During inference, the model achieves high throughput: on the aforementioned NVIDIA A800 GPU, it

processes a batch of 128 samples in an average of 127.39 ms, corresponding to just 0.9952 ms per sample.

During each training epoch, the model processes batches of data and computes the loss using the MSE criterion. The parameters are then updated through backpropagation. We continuously monitor the training MSE throughout this process, providing real-time insights into the model's learning progress. After each epoch, the model's performance is evaluated on the validation set, calculating the RMSE and Rp to assess the accuracy and correlation of the predicted binding affinities. The best-performing model, based on validation RMSE, is saved for use. In Table S2, we list the finally optimal hyperparameters used in the model.

## 5. RESULTS AND DISCUSSION

**5.1. Cutoff Distance Optimization for Pocket-Ligand Boundary Definition.** To determine the optimal spatial range for including protein pocket atoms around the ligand, we performed a sensitivity analysis on three cutoff distances (4 Å, 5 Å, and 6 Å) while keeping all other model parameters consistent. The performance metrics across these thresholds are summarized in Table 2, with the 5 Å configuration emerging as the optimal choice.

**Table 2. Evaluation of Model Performance At Different Thresholds**

| Metric | Threshold | Run 1 | Run 2 | Run 3 | Mean  | SD    |
|--------|-----------|-------|-------|-------|-------|-------|
| RMSE   | 4 Å       | 1.172 | 1.198 | 1.184 | 1.185 | 0.013 |
|        | 5 Å       | 1.162 | 1.155 | 1.155 | 1.157 | 0.004 |
|        | 6 Å       | 1.153 | 1.150 | 1.209 | 1.171 | 0.033 |
| Rp     | 4 Å       | 0.778 | 0.767 | 0.773 | 0.773 | 0.006 |
|        | 5 Å       | 0.783 | 0.785 | 0.785 | 0.784 | 0.002 |
|        | 6 Å       | 0.787 | 0.790 | 0.765 | 0.781 | 0.014 |

Empirically, the 4 Å cutoff yielded the weakest predictive performance ( $\text{RMSE} = 1.185 \pm 0.013$ ,  $\text{Rp} = 0.773 \pm 0.006$ ), likely due to its inability to capture long-range noncovalent interactions essential for binding (e.g., hydrophobic contacts extending up to 5 Å). In contrast, the 6 Å threshold introduced greater variance ( $\text{SD} = 0.014$  in Rp vs 0.002 at 5 Å) and increased computational burden, despite marginal improvements in individual runs. The 5 Å cutoff balanced accuracy and stability, achieving the lowest RMSE ( $1.157 \pm 0.004$ ) and highest Rp ( $0.784 \pm 0.002$ ) with minimal fluctuation across replicates.

This selection aligns with biophysical principles: 5 Å effectively encompasses key interaction ranges, including hydrogen bonds (2.5–3.5 Å), van der Waals forces (3–4 Å), and hydrophobic contacts (3–5 Å).<sup>61–65</sup> It also aligns with established practices in structure-based modeling, where 5 Å is widely adopted to capture essential binding interfaces without including irrelevant solvent-exposed atoms.<sup>10,66–68</sup>

**5.2. Meta-Path Selection for Graph Construction.** To enhance feature representation, we evaluated meta-path designs for constructing graphs from the original heterogeneous graph. Two candidate meta-paths were compared: short paths (“protein atom–ligand atom–protein atom” [P–L–P] and “ligand atom–protein atom–ligand atom” [L–P–L]) versus extended paths (“protein atom–ligand atom–ligand atom–protein atom” [P–L–L–P] and “ligand atom–protein atom–protein atom–ligand atom” [L–P–P–L]).

As shown in Table 3, the PLP/LPL paths outperformed extended paths in stability, with significantly lower variance in

**Table 3. Evaluation of Model Performance across Different Meta-Paths**

| Meta-path         | Metric | Run 1 | Run 2 | Run 3 | Average | SD    |
|-------------------|--------|-------|-------|-------|---------|-------|
| L–P–P–L & P–L–L–P | RMSE   | 1.181 | 1.159 | 1.144 | 1.161   | 0.019 |
|                   | Rp     | 0.776 | 0.785 | 0.791 | 0.784   | 0.007 |
| L–P–L & P–L–P     | RMSE   | 1.162 | 1.155 | 1.155 | 1.157   | 0.004 |
|                   | Rp     | 0.783 | 0.785 | 0.785 | 0.784   | 0.002 |

RMSE ( $\text{SD} = 0.004$  vs 0.019) and Rp ( $\text{SD} = 0.002$  vs 0.007). While extended paths occasionally showed marginal gains in individual runs (e.g., Run 3 RMSE = 1.144 for PLLP/LPPL), their performance was inconsistent due to noise amplification and redundant interaction patterns in longer paths, which obscured critical binding signals.<sup>69–71</sup>

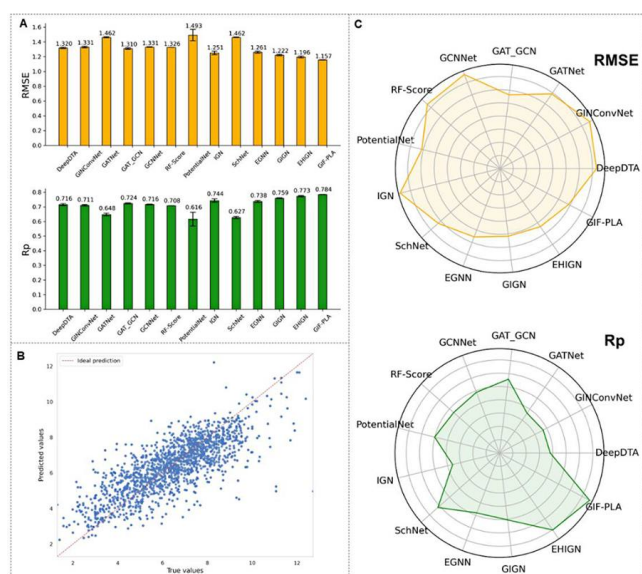
Notably, PLP/LPL paths retained equivalent predictive accuracy ( $\Delta\text{RMSE} = -0.004$ ,  $\Delta\text{Rp} = +0.002$ ) while preserving first-order proximity relationships—direct indicators of physical binding interactions. Their shorter length ensured computational efficiency and minimized extraneous information aggregation, striking an optimal balance between interaction fidelity and model robustness. These findings validate the utility of meta-paths in distilling essential features from heterogeneous graphs for binding affinity prediction.

**5.3. Performance Comparison with State-of-the-Art on Binding Affinity Prediction.** The performance of GIF-PLA was rigorously evaluated using benchmarking test set. GIF-PLA was compared against two complex-free models: DeepDTA and GraphDTA (including GCN, GAT, GIN, and GAT-GCN), as well as six complex-based models: RF-Score, IGN, SchNet, EGNN, GIGN, and EHIGN (See section 4.3.). All baseline models were implemented using the source code provided in their original publications and evaluated on the same training, validation, and test sets. Each experiment was conducted three times with different random seeds, and the final performance metrics were reported as the average and standard deviation of these three independent runs, as shown in Figure 2A. Tables S3 and S4 summarize the quantitative results.

The results demonstrate that GIF-PLA achieved state-of-the-art performance, exhibiting the lowest RMSE of  $1.157 \pm 0.004$  and a maximum Rp of  $0.784 \pm 0.002$ . Notably, GIF-PLA outperformed the nearest complex-based competitor (EHIGN) by 1.4% in Rp ( $\Delta = 0.011$ ). The data points closely clustered along the identity line (corresponding to ideal predictive performance) in Figure 2B demonstrate the model's effectiveness and validity. This superior performance of GIF-PLA may be attributed to its unique design that integrates structural topology and sequence patterns. This capability appears to be lacking in models that depend solely on either geometric structures or sequence information.

**5.4. Performance Evaluation on an External Test Set.** The model's predictive power was further challenged using the Test21 external validation set, which was designed to control protein and ligand similarity with the training data and is derived from non-PDBbind data sets (see Figure 2C and Supplementary Table S5 for results). Despite this stringent challenge, GIF-PLA maintained strong performance ( $\text{Rp} = 0.794$ ,  $\text{RMSE} = 1.174$ ). The Rp value of GIF-PLA surpasses that of other models, such as EHIGN and EGNN, which have Rp values of 0.763 and 0.722, respectively. This finding suggests that GIF-PLA relies less on





**Figure 2.** Performance evaluation. (A) RMSE and Rp results for GIF-PLA and baseline models on the test set. (B) Scatter plot illustrating the predicted binding affinities by GIF-PLA ( $y$ -axis) compared to experimentally measured affinities ( $x$ -axis) for the test set. (C) Performance evaluation for the models on the Test21 set.

memorizing training data and may suggest its ability to learn generalizable features.

**5.5. Performance Evaluation in the Clustering-Based Data Splitting.** In real-world drug discovery, PLA prediction models are frequently used to identify novel drugs or targets that are structurally distinct from those in the training data.<sup>55</sup> To mimic this scenario, clustering-based data splitting was employed for both ligands and proteins, as detailed in Section 4.2. This approach evaluates the model's ability to generalize to unseen chemical space.<sup>72</sup>

We performed three independent experiments for each of the three clustering-based data splitting strategies namely the ligand-clustering setting, protein-clustering setting, and both-clustering setting. The detailed results for each experiment are provided in Supplementary Tables S6–S8, with the average results summarized in Table 4. In the unseen Ligand scenario, our model improved average RMSE and Rp by 0.060 and 0.019 over

the second-best method EHIGN. For the unseen protein scenario, GIF-PLA achieves an average RMSE of 1.285 and an Rp of 0.752. In the most challenging all unseen scenario, our model shows improvements of 0.018 in RMSE and 0.08 in Rp compared to the second-best model. These results highlight the superior performance of our model relative to other models under more realistic generalization testing conditions.

**5.6. Ablation Study of Main Strategies.** We conducted ablation studies on GIF-PLA to explore the effects of different graph types and fusion methods. Under the same experimental setup, we implemented four simplified variants of GIF-PLA on the benchmark data set to understand the contribution of each component and assess their individual impact on the model's performance.

1. GIF-PLA $_{\alpha}$ : by only retaining the complex heterogeneous graph.
2. GIF-PLA $_{\beta}$ : by only retaining the complex heterogeneous graph and the homogeneous graphs, which are constructed based on the meta-paths.
3. GIF-PLA $_{\gamma}$ : by modifying the standard concatenation approach of the two homogeneous graphs to a gated fusion method.
4. GIF-PLA $_{\delta}$ : by only retaining the sequence modal data for proteins and ligands.

For each model, we performed three runs, recording the average values and standard deviations in Table 5, with individual run results detailed in Table S9. The results demonstrate that the GIF-PLA model, which simultaneously considers both sequence and structure modalities and employs dynamic weights for the two homogeneous graphs, demonstrates the best performance among all architectures.

The GIF-PLA $_{\alpha}$  variant, retaining only the complex heterogeneous graph architecture, exhibited substantial predictive capability (RMSE = 1.240, Rp = 0.752), confirming the fundamental importance of heterogeneous graph representations for PLA tasks. This performance stems from the graph's capacity to encode multimodal node features (atomic properties, bond types) and edge interactions, enabling comprehensive modeling of molecular topology through graph convolutional operations.

Incorporating meta-path enhanced architectures (GIF-PLA $_{\beta}$ ) yielded significant improvements, reducing RMSE by

**Table 4.** RMSE and Rp Values of GIF-PLA and Baseline Models under the Different Clustering-Based Data Splitting Settings<sup>a</sup>

| Model        | Ligand-clustering setting |                      | Protein-clustering setting |                      | Both-clustering setting |                      |
|--------------|---------------------------|----------------------|----------------------------|----------------------|-------------------------|----------------------|
|              | RMSE                      | Rp                   | RMSE                       | Rp                   | RMSE                    | Rp                   |
| DeepDTA      | 1.463 (0.006)             | 0.690 (0.004)        | 1.324 (0.052)              | 0.728 (0.028)        | 1.343 (0.081)           | 0.692 (0.047)        |
| GINConvNet   | 1.573 (0.019)             | 0.621 (0.021)        | 1.329 (0.011)              | 0.712 (0.007)        | 1.482 (0.024)           | 0.627 (0.020)        |
| GATNet       | 1.653 (0.012)             | 0.568 (0.006)        | 1.477 (0.021)              | 0.630 (0.012)        | 1.572 (0.085)           | 0.578 (0.023)        |
| GAT_GCN      | 1.666 (0.017)             | 0.561 (0.010)        | 1.327 (0.014)              | 0.722 (0.005)        | 1.450 (0.055)           | 0.622 (0.040)        |
| GCNNet       | 1.667 (0.014)             | 0.559 (0.012)        | 1.323 (0.008)              | 0.718 (0.002)        | 1.485 (0.027)           | 0.609 (0.017)        |
| RF-Score     | 1.500 (0.003)             | 0.660 (0.002)        | 1.385 (0.001)              | 0.700 (0.001)        | 1.290 (0.002)           | 0.713 (0.001)        |
| PotentialNet | 1.660 (0.049)             | 0.598 (0.047)        | 1.612 (0.064)              | 0.580 (0.005)        | 1.516 (0.022)           | 0.619 (0.026)        |
| IGN          | 1.397 (0.050)             | 0.728 (0.015)        | 1.393 (0.016)              | 0.688 (0.009)        | 1.316 (0.054)           | 0.696 (0.031)        |
| SchNet       | 1.656 (0.033)             | 0.573 (0.012)        | 1.537 (0.015)              | 0.604 (0.010)        | 1.426 (0.024)           | 0.643 (0.010)        |
| EGNN         | 1.366 (0.017)             | 0.731 (0.007)        | 1.382 (0.023)              | 0.699 (0.007)        | 1.317 (0.062)           | 0.712 (0.027)        |
| GIGN         | 1.374 (0.020)             | 0.736 (0.007)        | 1.363 (0.042)              | 0.717 (0.019)        | 1.300 (0.078)           | 0.708 (0.042)        |
| EHIGN        | 1.354 (0.008)             | 0.744 (0.005)        | 1.303 (0.019)              | 0.745 (0.005)        | 1.255 (0.019)           | 0.735 (0.007)        |
| GIF-PLA      | <b>1.294</b> (0.010)      | <b>0.763</b> (0.004) | <b>1.285</b> (0.009)       | <b>0.752</b> (0.010) | <b>1.237</b> (0.035)    | <b>0.743</b> (0.012) |

<sup>a</sup> **Bold** indicates the best result.

Table 5. Performance Comparison of GIF-PLA Variants in Ablation Study on the Test Set<sup>a</sup>

| Model     | Heterogeneous Graph | Meta-paths | Sequences | Gated fusion | RMSE         |       | Rp           |       |
|-----------|---------------------|------------|-----------|--------------|--------------|-------|--------------|-------|
|           |                     |            |           |              | Mean         | SD    | Mean         | SD    |
| GIF-PLA_α | ✓                   | ×          | ×         | ×            | 1.240        | 0.032 | 0.752        | 0.012 |
| GIF-PLA_β | ✓                   | ✓          | ×         | ×            | 1.183        | 0.011 | 0.774        | 0.005 |
| GIF-PLA_γ | ✓                   | ✓          | ×         | ✓            | 1.179        | 0.007 | 0.776        | 0.003 |
| GIF-PLA_δ | ×                   | ×          | ✓         | ×            | 1.324        | 0.003 | 0.713        | 0.008 |
| GIF-PLA   | ✓                   | ✓          | ✓         | ✓            | <b>1.157</b> | 0.004 | <b>0.784</b> | 0.002 |

<sup>a</sup> **Bold** indicates the best result.

0.057 and increasing Rp by 0.022. This enhancement demonstrates meta-paths' critical role in resolving long-range dependency limitations - a fundamental constraint in conventional graph networks. Specifically, meta-paths establish indirect connectivity between nonadjacent atoms through shared intermediaries, effectively modeling multihop chemical interactions essential for accurate affinity prediction.

For the third model variant GIF-PLA\_γ, we implemented a dynamic fusion approach that replaced the standard concatenation between the two homogeneous graphs, resulting in a marginal improvement in performance. Although this enhancement was not statistically significant, it contributed to the model's stability, as evidenced by the reduced standard deviation of the results. This increased stability is particularly beneficial in practical applications, as it minimizes the model's variability when confronted with new data, thereby enhancing its reliability.

The variant GIF-PLA\_δ, which utilizes only sequence modality, exhibits the lowest performance among the models, yet it still demonstrates a certain level of predictive capability. GIF-PLA\_δ may not achieve the same level of accuracy as its counterparts that utilize structure modal, it has the potential to provide crucial biological functional information within the sequence.

In summary, these results highlight the importance of integrating both sequence and structure information for accurate binding affinity prediction. The ablation studies also demonstrate the importance of meta-paths for capturing long-range dependencies between atoms and the potential benefits of dynamic fusion methods for improving model stability. From the biological perspective, heterogeneous graphs encode atomic and bonding structures as the structural foundation of binding, meta-paths capture long-range chemical dependencies akin to cooperative and allosteric effects, gated fusion reflects the adaptive nature of molecular recognition, and sequence modality carries evolutionary functional signals. These interpretations further justify the necessity of integrating both structural and functional perspectives in protein–ligand affinity modeling.

**5.7. Analysis of Model Interpretability in Ligand Pose Differentiation.** A systematic quantitative assessment of pose differentiation capabilities was performed using the CASF-2016 benchmark data set,<sup>73</sup> a widely recognized resource for evaluating protein–ligand complex prediction methods. The benchmark contains approximately 100 generated docking poses per complex with RMSD values spanning 0–10 Å, providing a rigorous testbed for distinguishing native-like conformations (defined as RMSD < 2 Å) from decoy structures.

Following CASF-2016 evaluation protocols, model performance was quantified through success rate metrics calculated as (Number of successful recognitions/Sample size). A successful identification was defined as the presence of at least one native-like pose within the top-ranked *N* predictions (*N* = 1, 2, 3). As

summarized in Table 6, GIF-PLA demonstrated notable generalization capability in discriminating relevant poses despite

Table 6. Pose Differentiating Power in Terms of Success Rate (%) between the Predicted Binding Affinity and RMSD Value for Different Complex-based Methods<sup>a</sup>

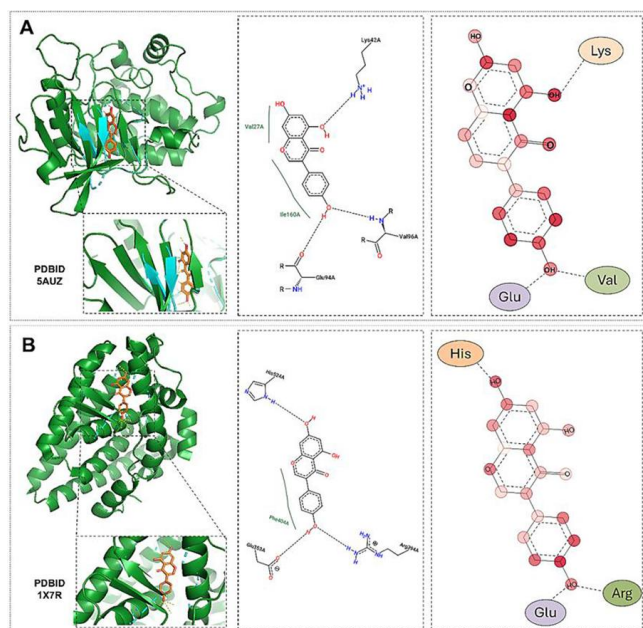
| Model        | Success rate <sup>b</sup> (%) |              |              | Success rate <sup>c</sup> (%) |              |              |
|--------------|-------------------------------|--------------|--------------|-------------------------------|--------------|--------------|
|              | Top 1                         | Top 2        | Top 3        | Top 1                         | Top 2        | Top 3        |
| PotentialNet | 19.65                         | 43.16        | 53.68        | 19.30                         | 42.11        | 52.28        |
| IGN          | 26.67                         | 42.81        | 49.82        | 26.32                         | 42.11        | 49.82        |
| SchNet       | 34.74                         | 49.12        | 57.89        | 34.74                         | 48.77        | 57.89        |
| EGNN         | 39.30                         | <b>58.25</b> | <b>66.67</b> | 38.25                         | <b>57.19</b> | <b>64.91</b> |
| EHIGN        | 41.40                         | 53.68        | 61.40        | <b>41.40</b>                  | 53.68        | 60.70        |
| GIF-PLA      | <b>43.50</b>                  | 57.60        | 63.70        | 40.50                         | 55.30        | 61.50        |

<sup>a</sup> **Bold** indicates the best result. Underlined results indicate the second-best result. <sup>b</sup>With native pose. <sup>c</sup>Without native pose.

never encountering decoy samples during training. Among complex-based models, GIF-PLA ranks among the top performers, achieving the highest or second-highest success rates.

While this demonstrated pose differentiation capability shows promise, it should be noted that current implementations primarily optimize for binding affinity prediction rather than explicit pose discrimination. Strategic enhancements such as incorporating physics-based energy terms or expanding training data sets with carefully curated decoy poses [16] could further improve this capability.

To complement our quantitative analysis, we conducted an in-depth case study examining GIF-PLA's performance on apigenin binding to two distinct protein targets (PDB code: SAUZ and 1 × 7R). The results are illustrated in Figure 3A,B. On the left, the structural overview provides a comprehensive view of the complex and a three-dimensional pocket perspective. The protein is shown in green, the ligand in orange, and the surrounding protein atoms within 5 Å in blue. This visualization was achieved using PyMOL software (PyMOL | [pymol.org](http://pymol.org)). The central part of the figure shows the two-dimensional interaction, with solid lines representing covalent bonds and dashed lines indicating noncovalent bonds. The 2D diagram is based on the 3D structure of the protein–ligand complex and was created using the PoseView function of ProteinPlus (<https://proteins.plus/>).<sup>74</sup> On the right side of the figure, the prediction advantages of GIF-PLA are clearly presented. The predicted values are prominently highlighted in red, with deeper colors indicating higher importance assigned by the model to the respective atoms. From the figure, it is evident that our model has learned to assign varying levels of importance to different atoms within the same ligand across different binding poses. This nuanced understanding enhances our ability to interpret the interaction between the protein and the ligand.

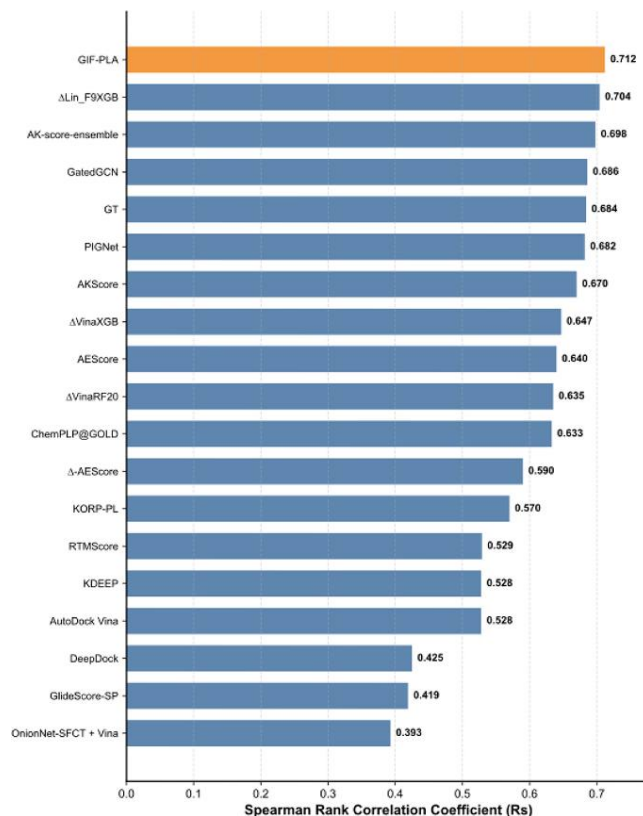


**Figure 3.** Binding-mode visualizations of the protein–ligand complexes with PDB IDs 5AUZ (A) and 1X7R (B). This figure presents the binding interactions in three panels: the left panel shows a 3D structural overview generated using PyMOL, where the protein is colored green, the ligand orange, and surrounding protein atoms within 5 Å in blue; the center panel displays a 2D interaction diagram created with PoseView, with solid lines representing covalent bonds and dashed lines indicating noncovalent interactions; the right panel illustrates the prediction outputs of the GIF-PLA model, where ligand atoms are color-coded in red, and deeper red hues reflect higher importance as assigned by the model.

**5.8. Evaluation of Ranking Power.** To evaluate the ability of our model to prioritize ligands by binding affinity for individual target proteins, a key requirement in virtual screening, we assessed its ranking power using the CASF-2016 benchmark.<sup>73</sup> Following the official protocol, ranking power is quantified as the average Spearman's rank correlation coefficient (Rs) between predicted and experimentally measured binding affinities across the 57 protein targets in the test set.

As shown in Figure 4, the proposed method, GIF-PLA, achieves an Rs of 0.712, demonstrating robust ranking performance. This result places GIF-PLA among the top-performing scoring functions currently available. Specifically, our model significantly outperforms several established approaches, including AutoDock Vina<sup>4</sup> (Rs = 0.528), GlideScore-SP<sup>75</sup> (0.419), KDEEP<sup>76</sup> (0.528), and DeepDock<sup>77</sup> (0.425), as well as more recent machine learning methods such as  $\Delta$ VinaRF20<sup>78</sup> (0.635),  $\Delta$ VinaXGB<sup>79</sup> (0.647),  $\Delta$ -AEScore<sup>80</sup> (0.590), and  $\Delta$ Lin\_F9XGB<sup>81</sup> (0.704). Moreover, GIF-PLA achieves performance that is either superior or comparable to state-of-the-art deep learning models, including PIGNet<sup>82</sup> (0.682), GT<sup>83</sup> (0.684), GatedGCN<sup>84</sup> (0.686), AKScore<sup>85</sup> (0.670), and AK-score-ensemble<sup>85</sup> (0.698). Additional compared methods are summarized in Figure 4, such as ChemPLP@GOLD<sup>86</sup> (0.633), AEScore<sup>80</sup> (0.640), RTMScore<sup>87</sup> (0.529), OnionNet-SFCT + Vina<sup>88</sup> (0.393), and KORP-PL<sup>89</sup> (0.570).

These results indicate that our structure-aware heterogeneous information fusion strategy effectively captures the critical structural features governing relative binding affinity. The strong and consistent ranking performance of GIF-PLA across diverse protein targets highlights its excellent generalization capability



**Figure 4.** Comparison of ranking power on the CASF-2016 benchmark core set.

and practical utility in real-world drug discovery, particularly in lead identification and optimization, where accurate prioritization of candidate compounds is essential.

**5.9. Virtual Screening Performance across Diverse Targets.** To validate the practical utility of GIF-PLA in real-world drug discovery applications, we conducted systematic virtual screening evaluations using seven pharmaceutically relevant targets (FEN1, KAT2A, PKM2, ALDH1, GBA, MAPK1, and VDR) from the LIT-PCBA benchmark data set (<http://drugdesign.unistra.fr/LIT-PCBA>). Distinctively, our evaluation protocol consolidated the original training and validation sets into a unified assessment framework, intentionally omitting target-specific parameter optimization to rigorously test the model's inherent generalization capacity. This design philosophy contrasts with conventional complex-based methods, such as GIGN<sup>11</sup> and EHIGN<sup>10</sup> that require extensive retraining for each new screening task. By eliminating the need for target-specific retraining, this measure enhances operational efficiency and enables the validation of the model's applicability across various drug discovery pipelines.

We benchmarked GIF-PLA against three state-of-the-art complex-based models (GIGN,<sup>11</sup> EGNN,<sup>60</sup> and EHIGN<sup>10</sup>) using enrichment factor (EF) analysis. The EF metric quantifies the ratio of active compounds identified in the top-ranked predictions versus random selection,<sup>90</sup> with higher values indicating superior early recognition capability critical for lead optimization. Following established practices in virtual screening validation, we computed EF values at three clinically relevant thresholds: EF0.005 (top 0.5%), EF0.01 (top 1%), and EF0.05 (top 5%), enabling multiscale assessment of model performance under varying stringency conditions.



As detailed in Table 7, GIF-PLA demonstrates remarkable consistency in prioritizing active compounds across all targets

**Table 7. Enrichment Factors of GIF-PLA in Drug Screening for Various Proteins<sup>a</sup>**

| Protein | Model   | EF0.005     | EF0.01      | EF0.05      |
|---------|---------|-------------|-------------|-------------|
| MAPK1   | GIGN    | 0.65        | 1.64        | 1.31        |
|         | EGNN    | 1.96        | <b>2.29</b> | 1.51        |
|         | EHIGN   | 0.65        | 0.66        | 1.38        |
|         | GIF-PLA | <b>2.62</b> | 1.64        | <b>2.36</b> |
| GBA     | GIGN    | 1.23        | 1.23        | 1.84        |
|         | EGNN    | 1.23        | 1.23        | 1.60        |
|         | EHIGN   | <b>2.45</b> | <b>3.07</b> | <b>3.44</b> |
|         | GIF-PLA | <b>2.45</b> | 1.84        | 2.21        |
| FEN1    | GIGN    | 0           | 0           | 0.61        |
|         | EGNN    | 1.11        | 1.11        | 0.50        |
|         | EHIGN   | <b>1.67</b> | 2.23        | <b>1.78</b> |
|         | GIF-PLA | <b>1.67</b> | <b>2.51</b> | 1.28        |
| VDR     | GIGN    | <b>0.87</b> | 0.65        | 1.04        |
|         | EGNN    | 0.43        | 0.43        | 0.52        |
|         | EHIGN   | 0.65        | <b>0.98</b> | 0.39        |
|         | GIF-PLA | 0.65        | 0.81        | <b>1.21</b> |
| KAT2A   | GIGN    | 0           | 0           | <b>0.93</b> |
|         | EGNN    | 0           | 0           | 0.31        |
|         | EHIGN   | 0           | <b>1.04</b> | 0.73        |
|         | GIF-PLA | <b>1.04</b> | 0.52        | 0.83        |
| ALDH1   | GIGN    | <b>2.21</b> | 1.79        | 1.42        |
|         | EGNN    | 1.57        | 1.63        | 1.42        |
|         | EHIGN   | 2.13        | <b>2.02</b> | 1.45        |
|         | GIF-PLA | 2.09        | 1.76        | <b>1.70</b> |
| PKM2    | GIGN    | 0           | 0.37        | 0.92        |
|         | EGNN    | 0           | 0           | 0.61        |
|         | EHIGN   | 0           | 1.10        | 1.32        |
|         | GIF-PLA | <b>1.47</b> | <b>1.65</b> | 1.36        |

<sup>a</sup>Bold indicates the best result.

and EF thresholds. For instance, in PKM2 screening, the model achieved EF values of 1.47 (EF0.005), 1.65 (EF0.01), and 1.36 (EF0.05), maintaining robust performance regardless of selection stringency. Notably, while comparator models occasionally failed to identify actives (EF = 0), GIF-PLA consistently delivered positive enrichment factors across all experimental conditions.

## 6. CONCLUSIONS

In this work, we present GIF-PLA, a complex structure-based deep learning framework for PLA prediction. Our approach aims to enhance the generalization of PLA prediction by seamlessly integrating both heterogeneous protein–ligand graphs, derived from complex structural and sequence data. In addition to extracting topological interaction details from the heterogeneous graph, GIF-PLA strives to acquire a nuanced understanding of high-order relationships. This is achieved through the definition of meta-paths within the graph and the comprehensive integration of sequences with structures, enabling the prediction of PLA based on both graph-based and sequence-based information.

Compared with state-of-the-art methods, including both complex structure-free and complex structure-based models, experimental results demonstrate that GIF-PLA achieves higher accuracy and lower error in predicting PLA across several benchmark data sets. The model also exhibits optimal

performance on data with controlled sequence similarity as well as on new data sources, highlighting the generalizability and superior scoring capability of GIF-PLA. Visualizations of the learned graphs demonstrate the model's ability to discern key interaction atoms.

Overall, GIF-PLA is a novel tool for PLA prediction based on multimodal data. Its role extends beyond merely improving prediction accuracy; it stands as a potential source of invaluable insights for drug discovery. Despite our success, there is still room for improvement. First, the model's dependency on experimentally determined protein–ligand complex structures restricts its application in scenarios where such structural information is unavailable. Second, the predictions have not been validated through prospective wet-lab experiments, leaving their practical utility in real-world drug screening contexts to be further established.

To address these limitations, several promising research directions could be explored. Future work may focus on developing streamlined pipelines that integrate rapid structure prediction tools with affinity assessment to reduce reliance on experimental structures. Additionally, establishing a framework for iterative feedback between computational predictions and experimental validation would help build a more robust and continuously improving evaluation system. The integration of affinity prediction modules with generative models also presents an interesting opportunity to advance structure-informed compound design. Pursuing these directions could significantly enhance GIF-PLA's applicability and practical value, potentially transforming it into a more versatile tool for drug discovery.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

All data and source codes are available on GitHub (<https://github.com/yanzhu22/GIF-PLA>).

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.5c01927>.

Distribution of protein and SMILES sequence lengths (Figure S1); similarity matrices for ligands and proteins (Figure S2); detailed features for nodes and edges (Table S1); optimal hyperparameters (Table S2); external validation set results (Table S5); complete experimental results for ligand-based, protein-based, and dual clustering data splits (Tables S6–S8) (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Junjie Wang** – Department of Medical Informatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211166, China; [orcid.org/0000-0003-0732-051X](https://orcid.org/0000-0003-0732-051X); Email: [junjie2021@njmu.edu.cn](mailto:junjie2021@njmu.edu.cn)

**Lingling Zhao** – Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China; [orcid.org/0000-0002-2808-3771](https://orcid.org/0000-0002-2808-3771); Email: [zhaoll@hit.edu.cn](mailto:zhaoll@hit.edu.cn)

### Authors

**Yan Zhu** – Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China; Department of Health Technology and Informatics, Faculty of Health and Social Sciences, Hong Kong Polytechnic University, Hong Kong 999077, China; [orcid.org/0000-0001-8692-0817](https://orcid.org/0000-0001-8692-0817)

Chunyu Wang – Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China; [orcid.org/0000-0002-2965-9920](https://orcid.org/0000-0002-2965-9920)

Quan Zou – Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou 324003, China; [orcid.org/0000-0001-6406-1142](https://orcid.org/0000-0001-6406-1142)

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acs.jcim.5c01927>

## Author Contributions

C.W., J.W., and L.Z. conceived the initial idea and designed the methodology. J.W. and Y.Z. implemented the algorithm, conducted the experiments, and processed the results. Y.Z. wrote the manuscript. J.W., Q.Z., and Y.Z. revised the manuscript.

## Funding

This work has been supported by the National Natural Science Foundation of China (NSFC, Nos. 62171164, 62272136, 62231013, 62131004, 62102191, and 62271174).

## Notes

The authors declare no competing financial interest.

## REFERENCES

- (1) Rajkhowa, S.; Deka, R. C. Protein-ligand docking methodologies and its application in drug discovery. In *Oncology: Breakthroughs in Research and Practice*; IGI Global, 2017; pp 891–914.
- (2) Khan, T.; Lawrence, A. J.; Azad, I.; Raza, S.; Joshi, S.; Khan, A. R. Computational drug designing and prediction of important parameters using in silico methods-A review. *Current Computer-Aided Drug Design* **2019**, *15* (5), 384–397.
- (3) Lionta, E.; Spyrou, G.; K Vassilatis, D.; Cournia, Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Current topics in medicinal chemistry* **2014**, *14* (16), 1923–1938.
- (4) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **2010**, *31* (2), 455–461.
- (5) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein–ligand docking using GOLD. *Proteins: Struct., Funct., Bioinf.* **2003**, *52* (4), 609–623.
- (6) Öztürk, H.; Özgür, A.; Ozkirimli, E. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* **2018**, *34* (17), i821–i829.
- (7) Nguyen, T.; Le, H.; Quinn, T. P.; Nguyen, T.; Le, T. D.; Venkatesh, S. GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics* **2021**, *37* (8), 1140–1147.
- (8) Stepniewska-Dziubinska, M. M.; Zielenkiewicz, P.; Siedlecki, P. Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. *Bioinformatics* **2018**, *34* (21), 3666–3674.
- (9) Jiang, D.; Hsieh, C.-Y.; Wu, Z.; Kang, Y.; Wang, J.; Wang, E.; Liao, B.; Shen, C.; Xu, L.; Wu, J. Interactiongraphnet: A novel and efficient deep graph representation learning framework for accurate protein–ligand interaction predictions. *Journal of medicinal chemistry* **2021**, *64* (24), 18209–18232.
- (10) Yang, Z.; Zhong, W.; Lv, Q.; Dong, T.; Chen, G.; Chen, C. 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* **2024**, *46* (12), 8191–8208.
- (11) Yang, Z.; Zhong, W.; Lv, Q.; Dong, T.; Yu-Chian Chen, C. Geometric interaction graph neural network for predicting protein–ligand binding affinities from 3d structures (gign). *journal of physical chemistry letters* **2023**, *14* (8), 2020–2033.
- (12) Chang, Y.; Chen, C.; Hu, W.; Zheng, Z.; Zhou, X.; Chen, S. Megnn: Meta-path extracted graph neural network for heterogeneous graph representation learning. *Knowledge-Based Systems* **2022**, *235*, No. 107611.
- (13) Kong, X.; Yu, P.; Ding, Y.; Wild, D. Meta path-based collective classification in heterogeneous information networks. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, 2012; pp 1567–1571.
- (14) Meng, C.; Cheng, R.; Maniu, S.; Senellart, P.; Zhang, W. Discovering meta-paths in large heterogeneous information networks. In *Proceedings of the 24th international conference on world wide web*, 2015; pp 754–764.
- (15) Chandra, D.; Leopold, J.; Fu, Y. NodeSense2Vec: Spatiotemporal Context-Aware Network Embedding for Heterogeneous Urban Mobility Data. In *2021 IEEE International Conference on Big Data (Big Data)*, 2021; IEEE: pp 2884–2893.
- (16) Sun, Y.; Han, J. Meta-path-based search and mining in heterogeneous information networks. *Tsinghua Science and Technology* **2013**, *18* (4), 329–338.
- (17) Fu, T.-y.; Lee, W.-C.; Lei, Z. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017; pp 1797–1806.
- (18) Zhao, X.; Zhao, X.; Yin, M. Heterogeneous graph attention network based on meta-paths for lncrna–disease association prediction. *Briefings in bioinformatics* **2022**, *23* (1), bbab407.
- (19) Zhong, K.; Wen, M.; Meng, F.; Li, X.; Jiang, B.; Zeng, X.; Li, Y. MMDTA: A Multimodal Deep Model for Drug–Target Affinity with a Hybrid Fusion Strategy. *J. Chem. Inf Model* **2024**, *64* (7), 2878–2888.
- (20) Wang, P.; Zheng, S.; Jiang, Y.; Li, C.; Liu, J.; Wen, C.; Patronov, A.; Qian, D.; Chen, H.; Yang, Y. Structure-Aware Multimodal Deep Learning for Drug–Protein Interaction Prediction. *J. Chem. Inf Model* **2022**, *62* (5), 1308–1317.
- (21) Wu, H.; Liu, J.; Jiang, T.; Zou, Q.; Qi, S.; Cui, Z.; Tiwari, P.; Ding, Y. AttentionMGT-DTA: A multi-modal drug–target affinity prediction using graph transformer and attention mechanism. *Neural Networks* **2024**, *169*, 623–636.
- (22) Wang, G.; Zhang, H.; Shao, M.; Sun, S.; Cao, C. MMPD-DTA: Integrating Multi-Modal Deep Learning with Pocket-Drug Graphs for Drug–Target Binding Affinity Prediction. *J. Chem. Inf Model* **2025**, *65* (3), 1615–1630.
- (23) Kyro, G.; Smaldone, A.; Shee, Y.; Xu, C.; Batista, V. T-ALPHA: A Hierarchical Transformer-Based Deep Neural Network for Protein–Ligand Binding Affinity Prediction with Uncertainty-Aware Self-Learning for Protein-Specific Alignment. *J. Chem. Inf Model* **2025**, *65* (5), 2395–2415.
- (24) Zhang, Z.; He, X.; Long, D.; Luo, G.; Chen, S. Enhancing generalizability and performance in drug–target interaction identification by integrating pharmacophore and pre-trained models. *Bioinformatics* **2024**, *40* (Supplement\_1), i539–i547.
- (25) Özçelik, R.; Bağ, A.; Atil, B.; Barsbey, M.; Özgür, A.; Ozkirimli, E. A Framework for Improving the Generalizability of Drug–Target Affinity Prediction Models. *Journal of Computational Biology* **2023**, *30* (11), 1226–1239.
- (26) Öztürk, H.; Ozkirimli, E.; Özgür, A. WideDTA: prediction of drug–target binding affinity. *arXiv preprint arXiv:1902.04166* **2019**.
- (27) Li, T.; Zhao, X.-M.; Li, L. Co-VAE: Drug–target binding affinity prediction by co-regularized variational autoencoders. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2022**, *44* (12), 8861–8873.
- (28) Mukherjee, S.; Ghosh, M.; Basuchowdhuri, P. DeepGLSTM: deep graph convolutional network and LSTM based approach for predicting drug–target binding affinity. In *Proceedings of the 2022 SIAM international conference on data mining (SDM)*, 2022; SIAM: pp 729–737.
- (29) Zhijian, L.; Shaohua, J.; Yigao, L.; Min, G. GDGRU-DTA: predicting drug–target binding affinity based on GNN and double GRU. *arXiv preprint arXiv:2204.11857* **2022**.

- (30) Zhu, Z.; Yao, Z.; Zheng, X.; Qi, G.; Li, Y.; Mazur, N.; Gao, X.; Gong, Y.; Cong, B. Drug–target affinity prediction method based on multi-scale information interaction and graph optimization. *Computers in Biology and Medicine* **2023**, *167*, No. 107621.
- (31) Zhu, Z.; Zheng, X.; Qi, G.; Gong, Y.; Li, Y.; Mazur, N.; Cong, B.; Gao, X. Drug–target binding affinity prediction model based on multi-scale diffusion and interactive learning. *Expert Systems with Applications* **2024**, *255*, No. 124647.
- (32) Zhu, Z.; Ding, Y.; Qi, G.; Cong, B.; Li, Y.; Bai, L.; Gao, X. Drug–target affinity prediction using rotary encoding and information retention mechanisms. *Engineering Applications of Artificial Intelligence* **2025**, *147*, No. 110239.
- (33) Jiang, M.; Li, Z.; Zhang, S.; Wang, S.; Wang, X.; Yuan, Q.; Wei, Z. Drug–target affinity prediction using graph neural network and contact maps. *RSC Adv.* **2020**, *10* (35), 20701–20712.
- (34) Jiang, M.; Wang, S.; Zhang, S.; Zhou, W.; Zhang, Y.; Li, Z. Sequence-based drug-target affinity prediction using weighted graph neural networks. *BMC genomics* **2022**, *23* (1), 449.
- (35) Nguyen, T.; Nguyen, T.; Le, T.; Tran, T. Gefa: early fusion approach in drug-target affinity prediction. *IEEE/ACM transactions on computational biology and bioinformatics* **2022**, *19* (2), 718–728.
- (36) Ouyang, X.; Handoko, S.; Kwok, C. CSCORE: A SIMPLE YET EFFECTIVE SCORING FUNCTION FOR PROTEIN-LIGAND BINDING AFFINITY PREDICTION USING MODIFIED CMAC LEARNING ARCHITECTURE. *Journal of Bioinformatics and Computational Biology* **2011**, *9*, 1–14.
- (37) Ballester, P. J.; Mitchell, J. B. O. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics* **2010**, *26* (9), 1169–1175.
- (38) Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf Model* **2009**, *49* (4), 1079–1093.
- (39) Dunbar, J. B.; Smith, R. D.; Yang, C. Y.; Ung, P. M. U.; Lexa, K. W.; Khazanov, N. A.; Stuckey, J. A.; Wang, S. M.; Carlson, H. A. CSAR Benchmark Exercise of 2010: Selection of the Protein-Ligand Complexes. *J. Chem. Inf Model* **2011**, *51* (9), 2036–2046.
- (40) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583.
- (41) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493.
- (42) Cang, Z.; Wei, G.-W. TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions. *PLoS computational biology* **2017**, *13* (7), No. e1005690.
- (43) Wang, Z.; Zheng, L.; Liu, Y.; Qu, Y.; Li, Y.; Zhao, M.; Mu, Y.; Li, W. OnionNet-2: A Convolutional Neural Network Model for Predicting Protein-Ligand Binding Affinity Based on Residue-Atom Contacting Shells. *Front Chem.* **2021**, *9*, No. 753002.
- (44) Prat, A.; Abdel Atty, H.; Bastas, O.; Kamuntavičius, G.; Paquet, T.; Norvaiša, P.; Gasparotto, P.; Tal, R. HydraScreen: A Generalizable Structure-Based Deep Learning Approach to Drug Discovery. *J. Chem. Inf Model* **2024**, *64* (15), 5817–5831.
- (45) Wang, Y.; Wu, S.; Duan, Y.; Huang, Y. A point cloud-based deep learning strategy for protein–ligand binding affinity prediction. *Briefings in bioinformatics* **2022**, *23* (1), bbab474.
- (46) Cang, Z.; Mu, L.; Wei, G. W. Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening. *PLoS Comput. Biol.* **2018**, *14* (1), No. e1005929.
- (47) Zhang, X.; Gao, H.; Wang, H.; Chen, Z.; Zhang, Z.; Chen, X.; Li, Y.; Qi, Y.; Wang, R. PLANET: A Multi-objective Graph Neural Network Model for Protein-Ligand Binding Affinity Prediction. *J. Chem. Inf Model* **2024**, *64* (7), 2205–2220.
- (48) Luo, D.; Liu, D.; Qu, X.; Dong, L.; Wang, B. Enhancing Generalizability in Protein-Ligand Binding Affinity Prediction with Multimodal Contrastive Learning. *J. Chem. Inf Model* **2024**, *64* (6), 1892–1906.
- (49) Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBbind database: methodologies and updates. *Journal of medicinal chemistry* **2005**, *48* (12), 4111–4119.
- (50) Consortium, U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47* (D1), D506–d515.
- (51) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B. PubChem 2019 update: improved access to chemical data. *Nucleic acids research* **2019**, *47* (D1), D1102–D1109.
- (52) Pecina, A.; Fanfrlík, J.; Lepšík, M.; Řezáč, J. SQM2. 20: Semiempirical quantum-mechanical scoring function yields DFT-quality protein–ligand binding affinity predictions in minutes. *Nat. Commun.* **2024**, *15* (1), 1127.
- (53) Bento, A. P.; Hersey, A.; Félix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L. J.; De Veij, M.; Leach, A. R. An open source chemical structure curation pipeline using RDKit. *Journal of Cheminformatics* **2020**, *12*, 1–16.
- (54) Li, W.; Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22* (13), 1658–1659.
- (55) Li, S.; Wan, F.; Shu, H.; Jiang, T.; Zhao, D.; Zeng, J. MONN: A Multi-objective Neural Network for Predicting Compound-Protein Interactions and Affinities. *Cell Syst* **2020**, *10* (4), 308.
- (56) Lei, Y.; Li, S.; Liu, Z.; Wan, F.; Tian, T.; Li, S.; Zhao, D.; Zeng, J. A deep-learning framework for multi-level peptide-protein interaction prediction. *Nat. Commun.* **2021**, *12* (1), 5465.
- (57) Zhao, M.; Lee, W.-P.; Garrison, E.; Marth, G. SSW library: an SIMD Smith-Waterman C/C++ library for use in genomic applications. *PLoS one* **2013**, *8* (12), No. e82138.
- (58) Gower, J.; Ross, G. Minimum spanning trees and single linkage cluster analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **1969**, *18* (1), 54–64.
- (59) Schütt, K.; Kindermans, P.-J.; Sauceda Felix, H.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems* **2017**, 30.
- (60) Satorras, V.; Hoogeboom, E.; Welling, M. E (n) equivariant graph neural networks. In *International conference on machine learning*, 2021; PMLR: pp 9323–9332.
- (61) Hwang, H.; Dey, F.; Petrey, D.; Honig, B. Structure-based prediction of ligand-protein interactions on a genome-wide scale. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114* (52), 13685–13690.
- (62) Gao, M.; Skolnick, J. The distribution of ligand-binding pockets around protein-protein interfaces suggests a general mechanism for pocket formation. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (10), 3784–3789.
- (63) de Freitas, R. F.; Schapira, M. A systematic analysis of atomic protein-ligand interactions in the PDB. *Medchemcomm* **2017**, *8* (10), 1970–1981.
- (64) Roy, A.; Zhang, Y. Recognizing Protein-Ligand Binding Sites by Global Structural Alignment and Local Geometry Refinement. *Structure* **2012**, *20* (6), 987–997.
- (65) Barratt, E.; Bingham, R.; Warner, D.; Laughton, C.; Phillips, S.; Homans, S. Van der Waals Interactions Dominate Ligand–Protein Association in a Protein Binding Site Occluded from Solvent Water. *J. Am. Chem. Soc.* **2005**, *127* (33), 11827–11834.
- (66) Li, S.; Zhou, J.; Xu, T.; Huang, L.; Wang, F.; Xiong, H.; Huang, W.; Dou, D.; Xiong, H. GLaNT: Protein-Ligand Binding Affinity Prediction via Geometry-Aware Interactive Graph Neural Network. *Ieee Transactions on Knowledge and Data Engineering* **2024**, *36* (5), 1991–2008.
- (67) Zhang, S.; Jin, Y.; Liu, T.; Wang, Q.; Zhang, Z.; Zhao, S.; Shan, B. SS-GNN: A Simple-Structured Graph Neural Network for Affinity Prediction. *Acs Omega* **2023**, *8* (25), 22496–22507.
- (68) Kim, Y.; Jeong, Y.; Kim, J.; Lee, E.; Kim, W.; Choi, I. MolNet: A Chemically Intuitive Graph Neural Network for Prediction of Molecular Properties. *Chemistry-an Asian Journal* **2022**, *17* (1616).



- (69) Li, X.; Ding, D.; Kao, B.; Sun, Y.; Mamoulis, N.; Ieee. Leveraging Meta-path Contexts for Classification in Heterogeneous Information Networks. In *37th IEEE International Conference on Data Engineering (IEEE ICDE)*; Electr Network, Apr 19–22, 2021; pp 912–923. .
- (70) Fu, X.; Zhang, J.; Men, Z.; King, I.; Assoc Comp, M. MAGNN: Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding. In *29th World Wide Web Conference (WWW)*, Taipei, TAIWAN, Apr 20–24, 2020; pp 2331–2341. .
- (71) Lai, P.; Dai, Q.; Lu, Y.; Wang, Z.; Chen, M.; Wang, C. MIGP: Metapath Integrated Graph Prompt Neural Network. *Neural Networks* **2024**, 179.
- (72) Qiao, F.; Peng, X.; Ieee Comp, S. Uncertainty-guided Model Generalization to Unseen Domains. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; Electr Network, Jun 19–25, 2021; pp 6786–6796. .
- (73) Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J. Chem. Inf Model* **2019**, 59 (2), 895–913.
- (74) Stierand, K.; Rarey, M. Drawing the PDB: Protein-Ligand Complexes in Two Dimensions. *ACS Med. Chem. Lett.* **2010**, 1 (9), 540–545.
- (75) Friesner, R.; Banks, J.; Murphy, R.; Halgren, T.; Klicic, J.; Mainz, D.; Repasky, M.; Knoll, E.; Shelley, M.; Perry, J.; et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, 47 (7), 1739–1749.
- (76) Jiménez, J.; Škalič, M.; Martínez-Rosell, G.; De Fabritiis, G. K(DEEP): Protein-Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. *J. Chem. Inf Model* **2018**, 58 (2), 287–296.
- (77) Méndez-Lucio, O.; Ahmad, M.; del Rio-Chanona, E.; Wegner, J. A geometric deep learning approach to predict binding conformations of bioactive molecules. *Nature Machine Intelligence* **2021**, 3 (12), 1033–1039.
- (78) Wang, C.; Zhang, Y. Improving scoring-docking-screening powers of protein-ligand scoring functions using random forest. *J. Comput. Chem.* **2017**, 38 (3), 169–177.
- (79) Lu, J.; Hou, X.; Wang, C.; Zhang, Y. Incorporating Explicit Water Molecules and Ligand Conformation Stability in Machine-Learning Scoring Functions. *J. Chem. Inf Model* **2019**, 59 (11), 4540–4549.
- (80) Meli, R.; Anighoro, A.; Bodkin, M.; Morris, G.; Biggin, P. Learning protein-ligand binding affinity with atomic environment vectors. *J. Cheminform* **2021**, 13 (1), 59.
- (81) Yang, C.; Zhang, Y. Delta Machine Learning to Improve Scoring-Ranking-Screening Performances of Protein-Ligand Scoring Functions. *J. Chem. Inf Model* **2022**, 62 (11), 2696–2712.
- (82) Moon, S.; Zhung, W.; Yang, S.; Lim, J.; Kim, W. PIGNet: a physics-informed deep learning model toward generalized drug–target interaction predictions. *Chemical Science* **2022**, 13 (13), 3661–3673.
- (83) Dwivedi, V.; Bresson, X. A Generalization of Transformer Networks to Graphs. *ArXiv* **2020**, abs/2012.09699.
- (84) Dwivedi, V.; Joshi, C.; Laurent, T.; Bengio, Y.; Bresson, X. Benchmarking Graph Neural Networks. *ArXiv* **2023**, abs/2003.00982.
- (85) Kwon, Y.; Shin, W.; Ko, J.; Lee, J. AK-Score: Accurate Protein-Ligand Binding Affinity Prediction Using an Ensemble of 3D-Convolutional Neural Networks. *Int. J. Mol. Sci.* **2020**, 21, 8424.
- (86) Korb, O.; Stützel, T.; Exner, T. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J. Chem. Inf Model* **2009**, 49 (1), 84–96.
- (87) Shen, C.; Zhang, X.; Hsieh, C.-Y.; Deng, Y.; Wang, D.; Xu, L.; Wu, J.; Li, D.; Kang, Y.; Hou, T.; et al. A generalized protein–ligand scoring framework with balanced scoring, docking, ranking and screening powers. *Chemical Science* **2023**, 14 (30), 8129–8146.
- (88) Zheng, L.; Meng, J.; Jiang, K.; Lan, H.; Wang, Z.; Lin, M.; Li, W.; Guo, H.; Wei, Y.; Mu, Y. Improving protein-ligand docking and screening accuracies by incorporating a scoring function correction term. *Brief Bioinform* **2022**, 23 (3). .
- (89) Kadukova, M.; Machado, K.; Chacón, P.; Grudinin, S. KORP-PL: a coarse-grained knowledge-based scoring function for protein-ligand interactions. *Bioinformatics* **2021**, 37 (7), 943–950.
- (90) Truchon, J.-F.; Bayly, C. Evaluating Virtual Screening Methods: Good and Bad Metrics for the “Early Recognition” Problem. *J. Chem. Inf Model* **2007**, 47 (2), 488–508.



CAS BIOFINDER DISCOVERY PLATFORM™

## BRIDGE BIOLOGY AND CHEMISTRY FOR FASTER ANSWERS

Analyze target relationships,  
compound effects, and disease  
pathways

Explore the platform

**CAS**  
A Division of the  
American Chemical Society