Computational and Structural Biotechnology Journal 23 (2024) 4396–4405

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Research Article

MM-DRPNet: A multimodal dynamic radial partitioning network for enhanced protein–ligand binding affinity prediction

Dayan Liu, Tao Song, Shudong Wang *

College of Computer Science and Technology, China University of Petroleum (East China), Qingdao, 266580, Shandong, China

ARTICLE INFO

Dataset link: http://www.pdbbind.org.cn/download/CASF-2016.tar.gz

Keywords:
Multimodal fusion framework
Deep learning
Drug discovery
Dynamic radial partitioning

ABSTRACT

Accurate prediction of drug-target binding affinity remains a fundamental challenge in contemporary drug discovery. Despite significant advances in computational methods for protein-ligand binding affinity prediction, current approaches still face substantial limitations in prediction accuracy. Moreover, the prevalent methodologies often overlook critical three-dimensional (3D) structural information, thereby constraining their practical utility in computer-aided drug design (CADD). Here we present MM-DRPNet, a multimodal deep learning framework that enhances binding affinity prediction by integrating protein-ligand structural information with interaction features and physicochemical properties. The core innovation lies in our dynamic radial partitioning (DRP) algorithm, which adaptively segments 3D space based on complex-specific interaction patterns, surpassing traditional fixed partitioning methods in capturing spatial interactions. MM-DRPNet further incorporates molecular topological features to comprehensively model both structural and spatial relationships. Extensive evaluations on benchmark datasets demonstrate that MM-DRPNet significantly outperforms state-of-the-art methods across multiple metrics, with ablation studies confirming the substantial contribution of each architectural component. Source code for MM-DRPNet is freely available for download at https://github.com/Bigrock-dd/MMDRPv1.

1. Introduction

Accurately predicting the binding affinity between drugs and targets is a vital aspect of drug discovery [1]. The strength of the interaction between a drug molecule and its protein target directly affects its therapeutic effectiveness. Therefore, understanding and predicting these interactions is essential for the rational design of new medications. In recent years, various computational models have been developed to estimate protein-ligand binding affinity, offering the potential to streamline the drug discovery process by reducing the reliance on costly and time-consuming experimental tests [2,3]. Despite these advances, many current models struggle to achieve high accuracy, mainly due to difficulties in effectively capturing the three-dimensional structural details of protein-ligand interactions [4].

The complex spatial arrangement of atoms within both proteins and ligands, along with their dynamic interaction patterns, plays a crucial role in determining binding affinity [5]. Modern drug discovery has increasingly incorporated computational tools, such as virtual screening of large compound libraries and analysis of ligand binding modes, as integral components of the development pipeline [6]. They not only provide ligands that may bind to specific targets, but also explain observed phenomena, such as the conformational relationships of compounds. Thus, developing a model that accurately captures the intricate 3D interactions between proteins and ligands remains a significant challenge. Although various research approaches have yielded encouraging results, they have focused mainly on single-modal data and have neglected the types of interactions between proteins and molecules with different biological functions. Recent research has realized the limitations of single modal information (e.g. protein sequence or structure) and has begun to combine sequence and structure information for analysis [7–10]. However, many studies still ignore the heterogeneity between different modal information and the complexity of proteins and ligands in the process of specific binding [11][12].

To address these limitations, we propose MM-DRPNet, a novel multimodal framework that integrates three key innovations: (1) A Dynamic Radial Partitioning (DRP) method that adaptively captures spatial interaction features in protein-ligand complexes; (2) A comprehensive feature extraction strategy combining structural information, interaction features, and physicochemical properties; and (3) A multimodal fusion architecture that effectively aligns and integrates diverse types of

* Corresponding author.
E-mail address: wangsd@upc.edu.cn (S. Wang).

https://doi.org/10.1016/j.csbj.2024.11.050
Received 16 October 2024; Received in revised form 23 November 2024; Accepted 30 November 2024

Available online 4 December 2024

molecular information. Our framework consists of three core modules: protein-ligand interaction feature extraction, molecular structural feature extraction, and multimodal feature fusion. The interaction module employs DRP to process 3D structural information and physicochemical interactions, while the structural module utilizes Graph Attention Networks [13] to capture molecular topology. These features are then optimally combined through our fusion module for accurate affinity prediction.

2. Related work

Molecular docking is one of the main approaches to structure-based drug discovery [14]. In molecular docking, the conformation of favorable ligand-protein interactions can be determined, producing so-called ligand poses (poses, binding modes) [15]. Their interactions with proteins can be quantified by scoring functions. To achieve efficiency in molecular docking, ligand-protein interactions would be quantified by a simple scoring function. However, this simplified model of protein-ligand interactions is a major limitation of the method, manifested in particular in the inaccuracy of the ranking of poses and the poor performance in predicting absolute or relative binding free energies. In addition, the correct sampling of ligand binding patterns may be limited by induced fit effects as well as by the different conformational states of the proteins (proteins are often treated as rigid bodies), which lead to inaccuracies that outweigh the inaccuracies introduced by the scoring function itself [16]. In addition, sometimes intramolecular interactions are incorrectly modelled, such as the presence of distorted amide groups, structural impulses between ligands within proteins, or unrealistic three-dimensional (3D) structures [17]. And another key factor affecting molecular docking is the molecular mass of the input structure [18]. Despite the widespread use of molecular docking, its simplified scoring functions and rigid protein assumptions limit prediction accuracy.

Sequence-based protein-ligand interaction prediction methods leverage the amino acid sequence of proteins to infer binding affinity [19]. These methods typically extract relevant features from protein sequences, such as amino acid composition, physicochemical properties, and evolutionary information, to train machine learning or deep learning models [20,21]. Early approaches, such as support vector machines (SVM) and random forest models, have demonstrated promise by utilizing these sequence-derived features to predict protein-ligand interactions [22,23]. However, these models often struggle to capture the complex nonlinear relationships present in protein-ligand binding, limiting their predictive power [24][25][26]. While traditional machine learning methods showed initial promise, they struggled to capture complex nonlinear relationships in protein-ligand binding.

With the advent of deep learning, more sophisticated models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been employed to improve prediction accuracy by capturing higher-level sequence patterns and long-range dependencies within protein sequences [27–30]. CNNs are particularly adept at extracting local interaction patterns from sequences, while RNNs and their variants, such as long short-term memory (LSTM) networks, excel at modeling sequential data by accounting for the temporal relationships between amino acids in a protein chain [31–34]. Additionally, transformer-based models, which utilize self-attention mechanisms, have recently demonstrated superior performance in capturing long-range dependencies and complex interactions within protein sequences, further advancing the field of sequence-based prediction [35–38]. Some examples include, DeepDTA leverages CNN to predict drug-target binding affinities by extracting features directly from protein sequences and compound representations [21]. GraphDTA utilizes graph neural networks to model molecular structures as graphs, improving upon previous sequence-based methods like DeepDTA by more accurately capturing the complex topological interactions within compounds for drug-target binding affinity prediction [39]. TransformerCPI applies Transformer architectures to capture long-range dependencies in both protein sequences and compound representations, offering improvements over graph-based models like GraphDTA by enhancing the ability to model complex interactions with greater contextual understanding for compound-protein interaction prediction [40]. There are also some models for the 3D structure of complexes, such as the OnionNet [41], a CNN based on element-specific contacts between proteins and ligands dependent on distance and Pafnucy [42], a 3D CNN that employs some computer vision-derived strategies to encode the protein and the ligand. There is also KDeep [43], which employs 3D CNN to predict protein-ligand binding affinity. All of these models are, to some extent, characterized by a vectorized lattice within a ligand-centered cube representing the protein-ligand complex, and they demonstrate good performance in predicting protein-ligand binding [44–46]. Although deep learning approaches have significantly improved prediction accuracy, most existing methods still rely on single-modal information, overlooking the inherent multimodal nature of protein-ligand interactions.

3. Materials and methods

3.1. Data preparation

We utilized the PDBBind v2016 database [47] as our primary data source, which comprises high-quality three-dimensional structures and experimentally determined binding affinity data for 13,285 protein-ligand complexes. The database is hierarchically organized into the General Set, Refined Set, and Core Set (CASF-2013 [48] and CASF-2016 [49]). To ensure high-quality training data, we implemented rigorous filtering criteria on the General Set (9,228 complexes). Only X-ray crystal structures with resolution better than 2.5 Å were retained, and complexes with missing atoms or residues were excluded. After applying these quality control measures, 8,873 complexes qualified for our training dataset.

To prevent data leakage and ensure unbiased evaluation, we carefully partitioned the datasets. The 8,873 filtered complexes from the General Set were used as the training set. For the validation set, we first removed all complexes that overlapped with the CASF-2016 and CASF-2013 sets from the Refined Set (4,057 complexes), then randomly selected 1,000 complexes. For testing, we utilized all 285 complexes from the CASF-2016 core set as our primary test set and 108 complexes from the CASF-2013 core set as our secondary test set, ensuring no overlap with the training and validation sets to maintain evaluation integrity.

We established a systematic preprocessing pipeline to standardize all protein-ligand complexes. Initially, water molecules and non-essential ions were removed from the structures, and alternative conformations were eliminated. For ligand processing, structures in mol2 format were converted to PDB format, with chemical correctness verified and proper protonation states assigned at pH 7.4. The ligands were then docked with their corresponding receptor PDB files to maintain consistent binding poses. Following complex preparation, we determined the elemental type of each atom and calculated relevant atomic properties including partial charges and hydrophobicity. To standardize the measurement units, we transformed the binding affinity data into the negative logarithmic form according to:

$$pK_a = -\log_{10} K_x \tag{1}$$

where K_x represents $IC50$, K_i, or K_d.

The final step involved dynamic spatial distance feature extraction on the processed complexes. All preprocessing steps were automated using custom Python scripts to ensure reproducibility. To maintain consistency with previous research, no additional alterations were made to the protein-ligand complexes beyond these standard preprocessing steps.

3.2. Dynamic racial partitioning

To effectively model protein–ligand interactions, we use the atomic distances between the protein surface and the ligand as key descriptors. The 3D spatial arrangement of these atoms significantly influences binding affinity. To capture this spatial relationship, we divide the atomic pairs into multiple layers based on specific distance thresholds. Each layer includes pairs within a defined range, determined by cutoff values, which represent the maximum distances between protein and ligand atoms. By segmenting the interactions into these layers, we can analyze the interaction patterns across various spatial scales more effectively. $R = \{r_i\}_{i=1}^{N_r}$ represent the set of N_r atoms in the protein and $L = \{l_j\}_{j=1}^{N_l}$ represent the set of N_l atoms in the ligand. The spatial distance layer feature f_s is then defined as the number of atomic pairs within each layer:

$$f_s = \sum_{i=1}^{N_r} \sum_{j=1}^{N_l} \mathbb{I}\left(d_{\min} \le d_{ij} < d_{\max}\right) \tag{2}$$

$$d_{ij} = \left\| r_i - l_j \right\| \tag{3}$$

Here, d_{ij} represent distance between atom r_i in protein and atom l_j in ligand. $I(\cdot)$ is the indicator function, which equals 1 if $d_{min} \le d_{ij} < d_{max}$, and 0 otherwise. The parameters d_{min} and d_{max} represent the minimum and maximum distance thresholds for each feature layer, respectively.

The 3D structures and interaction patterns of proteins and ligands vary widely, and fixed distance partitioning may fail to capture all significant interactions. In particular, layer partitioning near distance boundaries can be overly coarse or imprecise, potentially missing critical details [50,45]. Therefore, a more flexible approach is required to ensure accurate modeling of these interactions in the next stages of our analysis.

We propose a dynamic radial partitioning (DRP) approach that adaptively adjusts the hierarchical granularity of 3D interaction modeling for each protein–ligand complex, effectively accommodating the inherent heterogeneity of the complexes. The method is flexible enough to capture key interactions. In regions with dense interactions (e.g., near the active site of a ligand), we use finer hierarchical partitioning to capture subtle changes in the interactions with higher precision; conversely, in regions with weak or distant interactions, we use coarser hierarchical partitioning to reduce unnecessary computational overheads and thus optimize the accuracy and efficiency of the analysis. Specifically, we first constructed a distance matrix $D = \{d_{ij}\}_{i=1,j=1}^{N_r,N_l}$ for all atom pairs between proteins and ligands. Then, this distance matrix was analyzed through clustering. We defined K spatial distance feature layers and clustered the distances using the K-means algorithm to achieve adaptive hierarchical classification. Experiments demonstrated that the DRP method substantially improved computational efficiency while maintaining high accuracy, and offered significant advantages over the traditional fixed-distance partitioning method.

$$L = \sum_{k=1}^{K} \sum_{i,j \in C_k} \left(d_{ij} - \mu_k\right)^2 \tag{4}$$

Here, C_k represents the set of atom pairs in the $K-th$ layer, and μ_k is the centroid distance of layer C_k. The objective of $K-means$ clustering is to minimize the squared error between the distances within each layer and their respective centroid. After clustering, each layer C_k contains a certain range of atom pairs. The dynamic radial partitioning feature f_s^{dynamic} is defined as the number of atom pairs within each layer.

In order to comprehensively capture the diverse range of interactions within protein-ligand complexes, we systematically defined various interaction types based on atomic composition and spatial geometries. These include well-characterized interactions such as hydrogen bonds, hydrophobic interactions, as well as more nuanced types like salt bridges and $\pi-\pi$ stacking. Specifically, hydrogen bonds are identified when the distance between donor and acceptor atoms is less than 3.5 Å, with an accompanying bond angle exceeding 120°, ensuring the geometrical specificity required for strong hydrogen bonding. Hydrophobic interactions, on the other hand, are defined as interactions between carbon atoms where the interatomic distance is less than 5.0 Å, reflecting the tendency of nonpolar residues to cluster and avoid aqueous environments.

Beyond these two primary interaction types, we also considered salt bridges, which form between oppositely charged residues, and $\pi-\pi$ stacking, a key interaction between aromatic rings, commonly seen in protein-ligand binding scenarios. For each radial layer surrounding the ligand, we computed the number of atomic pairs corresponding to each interaction type, effectively capturing the spatial distribution of these interactions within the protein-ligand complex. Furthermore, we quantified the frequencies of specific atomic pair combinations (e.g., C-C, C-N) for each interaction type, thereby incorporating finer granularity into the interaction modeling process.

This comprehensive approach led to the construction of a spatial distance feature vector of length 320 for each complex, effectively summarizing the complex network of interactions. By integrating both the type and frequency of atomic interactions within distinct spatial layers, we significantly enriched the model's ability to predict binding affinity by leveraging a multidimensional feature space that reflects the intricate molecular interactions driving protein-ligand binding.

Additionally, to capture different types of interactions, we defined various interaction types based on atomic types and spatial geometries, including hydrogen bonds, hydrophobic interactions and others such as salt bridges and $\pi-\pi$ stacking. Specifically, hydrogen bonds are defined as interactions with a donor-acceptor atom distance of less than 3.5 Å and a bond angle greater than 120°. Hydrophobic interactions are defined as those between carbon atoms with a distance less than 3.5 Å, and more details about criteria for identifying key interaction types are shown in Table S2. For each layer, we calculated the number of atom pairs for each interaction type, as well as the frequencies of different atomic pair combinations (e.g., C-C, C-N), resulting in a spatial distance feature vector of length 335.

3.3. Molecular structure feature extraction

In this study, we employed GAT, a specific type of Graph Neural Network (GNN), to effectively extract both topological and chemical features from the molecular structures of ligands. Initially, the SMILES strings of the ligands were converted into molecular graphs using RDKit, where nodes represent atoms and edges correspond to chemical bonds. The node features included various atomic properties, such as atom type (encoded via one-hot vectors), formal charge, hybridization state, and whether the atom belongs to a ring structure. Edge features were designed to capture bond characteristics, including bond type (e.g., single, double, aromatic), conjugation, and ring participation. We then constructed a GAT to update the node embeddings h_i through an attention-based mechanism. In this framework, nodes iteratively aggregate information from neighboring nodes, weighted by attention coefficients. By stacking multiple layers of GAT, each node integrates progressively more information from its local and extended molecular environment. This enables the model to capture both localized chemical interactions and more global structural features of the molecules, providing a comprehensive representation for downstream predictive tasks. Fig. 1B illustrates the process of extracting molecular structural features and the model updates node embeddings through self-attention mechanism according to:

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in N(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)}\right) \tag{5}$$

Where $N(i)$ represents the neighbors of node i, the updated node feature $h_i^{(l+1)}$ at layer $l+1$ for node i is obtained by aggregating the features

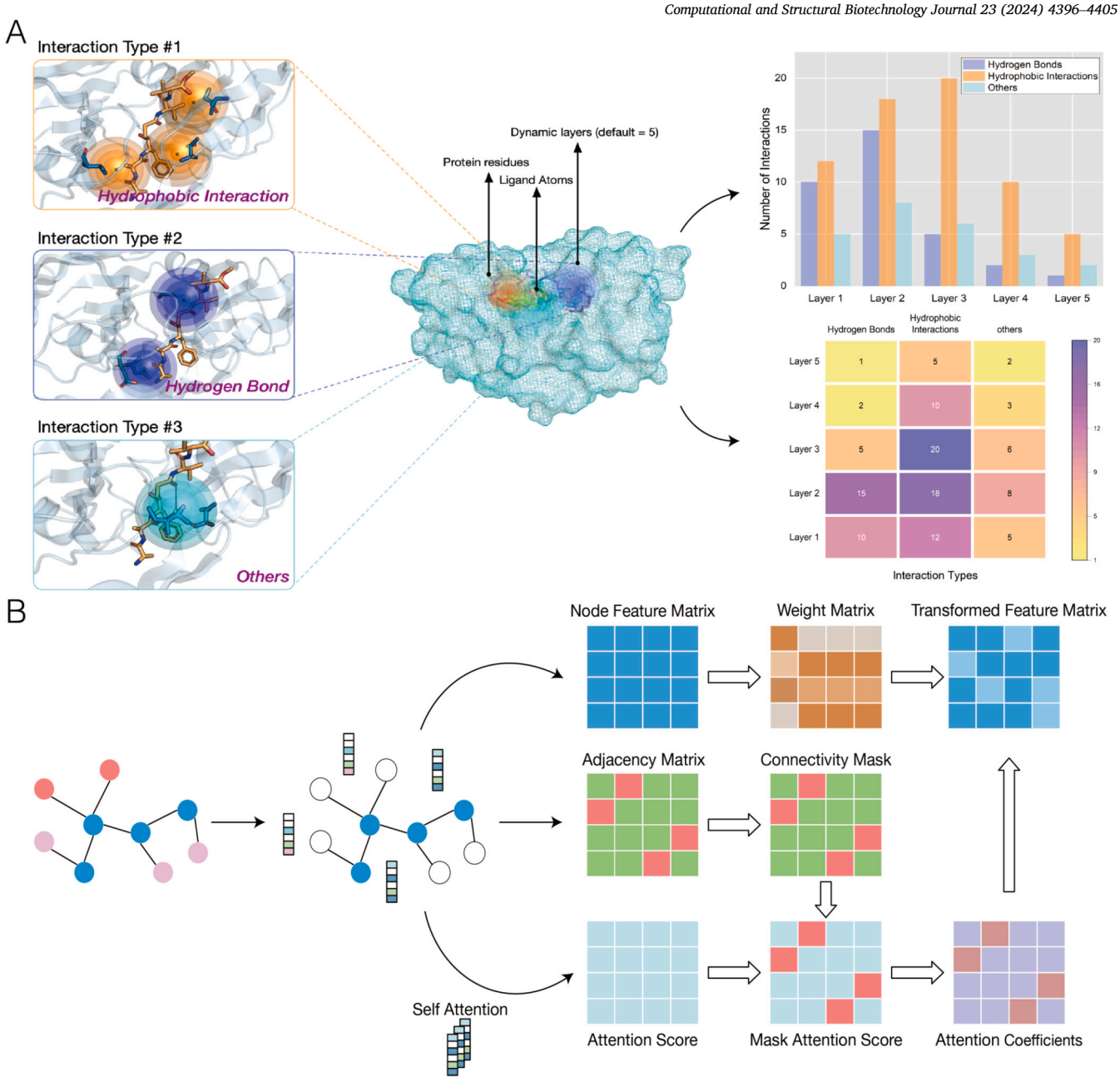

Fig. 1. Visualization of dynamic radial partitioning (DRP) and graph attention network (GAT) in protein-ligand interactions. (A) DRP partitions the 3D space around the ligand into dynamic layers, classifying interaction types such as hydrophobic interactions and hydrogen bonds, and showing the distribution across layers. (B) Schematic of the GAT architecture, highlighting how node features are updated via attention mechanisms, including operations on adjacency matrices, connectivity masks, and attention scores.

$h_j^{(l)}$ of its neighboring nodes $j \in N(i)$ from the previous layer. A learnable weight matrix $W^{(l)}$ is applied to transform these features, while the attention coefficient $\alpha_{ij}^{(l)}$ determines the importance of each neighboring node's contribution to node i. The weighted sum of these transformed features is then passed through a nonlinear activation function σ, resulting in the updated representation of node i. The final molecular graph features are obtained through a global pooling layer, resulting in a fixed-dimensional vector representation, which is then used for subsequent model training.

3.4. MM-DRPNet architecture

A dual-branch multimodal neural network architecture was constructed. The architecture of the model is illustrated in Fig. 2.

For each protein-ligand complex, the 3D interaction information is processed through DRP to generate matrices of different spatial regions, which are used for subsequent feature extraction. To effectively model the dynamic spatial layers of the protein-ligand complex, the model consists of three-layer convolutional layers capturing the local spatial interaction patterns between molecules, each followed by a CBAM (Convolutional Block Attention Module) to enhance the feature extraction capability of the model, followed by max pooling to reduce the sampling rate of the feature maps while retaining the most important features. After three layers of CNN + Max Pooling, a dense layer maps the extracted features to fixed-size interaction feature vectors.

Finally, the multimodal feature fusion module integrates features from two different modules: interaction features and molecular structure features extracted by the GAT. The interaction features capture the

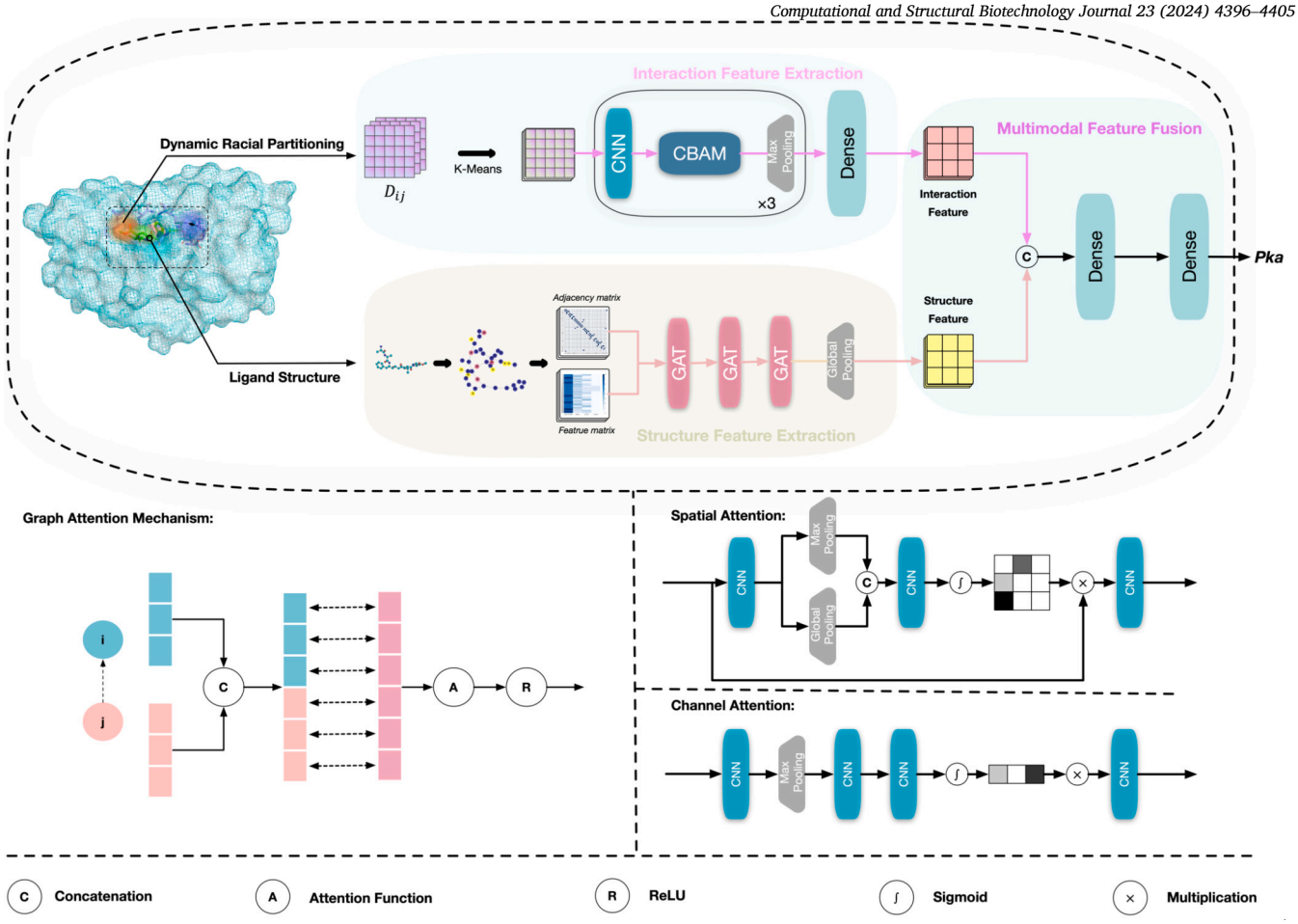

Fig. 2. Illustration of the MM-DRPNet framework.

spatial interaction information between the ligand and the target, while the structure features are derived from the topological information of the molecular structure. These features are concatenated and combined into a unified representation. In the fused feature vector, the model can simultaneously leverage both local molecular interactions and global structural information, providing more comprehensive support for subsequent prediction tasks. The fused features are then processed by fully connected layers to generate the final prediction result.

During model training, we implemented a custom-designed loss function that synergistically incorporates the root mean square error (RMSE) and the Pearson correlation coefficient (R), thereby optimizing both the predictive accuracy and the correlation between predicted and actual values. This dual-objective method ensures that the model not only minimizes error but also maximizes the consistency between the predicted and true binding affinities. The loss function is formally defined as:

$$\text{Loss} = \alpha(1 - R) + (1 - \alpha)\text{RMSE} \tag{6}$$

Inspired by previous research, we defined the weight parameter α as 0.8. R and RMSE were employed to balance the dual objectives of the model. The R measures the linear correlation between predicted and true values, while RMSE captures the magnitude of prediction errors.

We used the cosine annealing learning rate scheduling strategy, which dynamically adjusts the learning rate throughout training to mitigate the risk of the model converging to local optima. To further prevent overfitting, we incorporated dropout layers and applied L2 regularization. During training, we implemented an early stopping strategy, halting the process if the validation loss failed to decrease over five consecutive epochs and saving the model at its optimal state. See more details in the Supporting Information Table S1.

3.5. Evaluation metrics

In this study, we utilized two metrics to evaluate the error between the predicted and actual values. The Mean Absolute Error (MAE) measures the average absolute difference between the predicted and actual values, providing an indication of the degree of deviation in the model's predictions.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|pK_{a_{pred_i}} - pK_{a_{true_i}}| \tag{7}$$

In addition, The RMSE quantifies the relative deviation between the predicted and experimentally determined values by summing the squared residuals for each sample and then dividing by the total number of samples, and is more sensitive to large errors than MAE.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\text{p}K_{a_{\text{pred}}} - \text{p}K_{a_{\text{true}}}\right)^2} \tag{8}$$

The standard deviation (SD) as another metric, which was also adopted in the CASF-2013.

$$\text{SD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left((a^*\text{pK}_{a_{\text{pred}}} + b) - \text{pK}_{a_{\text{true}}}\right)^2} \tag{9}$$

where a and b represent the slope and intercept of the linear regression line fitted to the predicted and measured pK_a data points, respectively. These values provide key insights into the accuracy and consistency of the predictions, setting the stage for further evaluation in the subsequent analyses.

Finally, the correlation between the predicted and actual values is calculated using the Pearson correlation coefficient (R).

$$R = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})}} \quad (10)$$

where n is the sample size and x and y the predicted and the expected pK_a, respectively.

4. Results and discussion

4.1. Compare with SOTA DTA prediction models

Traditional machine learning methods usually rely on hand-designed features, which are difficult to capture complex nonlinear protein-ligand interactions and have limited performance in high-dimensional space. On the other hand, most deep learning methods mostly focus on unimodal information (e.g., sequence or structure), lack the ability to integrate features from different modalities, and are unable to adequately characterize complex interactions in three-dimensional space, limiting their accuracy and generalization ability in biomolecular prediction.

Our method addresses these limitations through several key innovations. The Dynamic Radial Partitioning (DRP) method uniquely captures spatial interaction features by adaptively adjusting to each protein-ligand complex's specific structural characteristics. The integration of Graph Attention Networks (GAT) enables precise modeling of molecular topology while preserving chemical properties. Additionally, our multimodal framework systematically combines spatial interactions (including hydrogen bonding and hydrophobic interactions) with molecular structural features, providing a comprehensive view of protein-ligand binding mechanisms.

We compared our model with several state-of-the-art DTA prediction methods, including traditional molecular docking approaches (such as AutoDock Vina), sequence-based models (e.g., PSICHIC), and structure-based deep learning models (e.g., OnionNet, Pafnucy). All models were trained and validated using identical dataset partitioning methods and evaluated on both the CASF-2016 and CASF-2013 core sets to ensure fair comparison. The comparative results are reported in Table 1. Our model achieved superior performance on the CASF-2016 core set with an RMSE of 1.128, MAE of 0.853, Pearson correlation coefficient of 0.884, and SD of 1.086, significantly outperforming all baselines.

A detailed comparison with recent approaches highlights the advantages of our model. The sequence-based PSICHIC model utilizes physicochemical graph neural networks to learn protein-ligand interaction fingerprints from sequence data. While this approach is computationally efficient without requiring 3D structural information, it cannot capture crucial spatial interaction patterns that are only observable in three-dimensional structures. In contrast, our DRP method directly captures these spatial relationships, leading to more accurate binding affinity predictions. The fixed layer partitioning strategy employed in OnionNet may miss crucial spatial information, whereas our DRP method dynamically adjusts the layer partitioning based on each complex's unique spatial structure. This adaptive approach enables flexible distance scaling that adapts to different binding pocket sizes, enhances the capture of local interaction patterns within each partition, and better preserves spatial relationship information between different regions.

The superiority of our dynamic approach is particularly evident in cases where binding sites have irregular shapes or varying sizes, where fixed partitioning methods might fail to capture important interaction features. Furthermore, by integrating molecular topological information through GAT, our model achieves more accurate predictions by considering both spatial and chemical properties simultaneously. This comprehensive approach explains the significant improvement in prediction accuracy observed in our experimental results.

Table 1
Performance comparison of different scoring functions on CASF-2013 and CASF-2016.[a]

Dataset	Model	MAE	RMSE	SD	R
CASF-2016	AutoDock Vina [51]	1.940	2.350	-	0.600
	DeepDTA [21]	1.148	1.443	1.445	0.749
	Pafncuy [42]	1.129	1.418	1.375	0.775
	PSICHIC [38]	1.040	1.336	-	0.792
	Onionnet [41]	0.980	1.278	1.260	0.816
	KDeep [43]	-	1.270	-	0.820
	DPLA [8]	0.972	1.255	1.248	0.820
	CurvAGN [7]	0.930	1.217	1.191	0.830
	DockingApp RF [51]	1.130	1.380	1.260	0.830
	AGL-Score [9]	-	1.271	-	0.833
	MFE [10]	0.882	1.151	1.138	0.851
	MM-DRPNet	**0.853**	**1.128**	**1.086**	**0.884**
CASF-2013	AutoDock Vina [48]	1.950	2.400	-	0.570
	DeepBindRG [51]	1.480	1.820	1.730	0.640
	Pafnucy [42]	1.510	1.620	1.610	0.700
	Onionnet [41]	1.210	1.500	1.450	0.780
	DockingApp RF [51]	1.130	1.380	1.260	0.790
	AGL-Score [9]	-	1.97	1.450	0.792
	MM-DRPNet	**0.915**	**1.212**	**1.198**	**0.831**

[a] These results are taken from their respective published papers;
Bold numbers represent the best performance in each metric column.

4.2. Impact of interaction type ablation on model performance

In constructing dynamic layer features, we focused on various interaction types between proteins and ligands, including hydrogen bonds, hydrophobic interactions, and others. These interactions play a crucial role in molecular recognition and binding processes. To assess the impact of different interaction types on the model's predictive performance, we designed a series of ablation experiments, where specific interaction features were systematically removed to observe the resulting changes in model performance. The results are shown in the Fig. 3A.

When hydrogen bond features were excluded, the RMSE increased to 1.239, and the R value dropped to 0.841, indicating the critical role of hydrogen bonds in maintaining model accuracy (The decay curves are shown in Supporting Information Figure S2). These bonds are essential in stabilizing protein-ligand complexes by creating specific geometric arrangements between molecules. The notable decline in performance upon their removal suggests that the model depends heavily on hydrogen bond features to capture essential aspects of molecular binding, consistent with their established importance in biological systems.

Removing hydrophobic interaction features led to an RMSE of 1.184 and a R of 0.866, showing a clear but less severe effect on model performance compared to hydrogen bonds. Hydrophobic interactions occur between non-polar regions of the molecule and help stabilize the binding of protein ligands by reducing contact with water, and removing these features weakens the model's ability to capture hydrophobicity drivers, leading to a decrease in predictive performance.

With the removal of other types of features, such as van der Waals forces, the model has an RMSE of 1.157 and a R of 0.874, which is a small change from the baseline performance. This may be due to the fact that van der Waals forces are weak and non-specific, resulting in the model being less sensitive to their removal and only a slight change in performance.

These results highlight the critical importance of hydrogen bond and hydrophobic interactions in the model's ability to predict protein-ligand binding accurately, which is consistent with their known roles in molecular biology. Van der Waals forces, while present, appear to play a less significant role in the model's performance. The ablation experiments provide clear evidence that the model effectively captures the most biologically relevant interaction types, reinforcing its robustness in predicting molecular binding affinities.

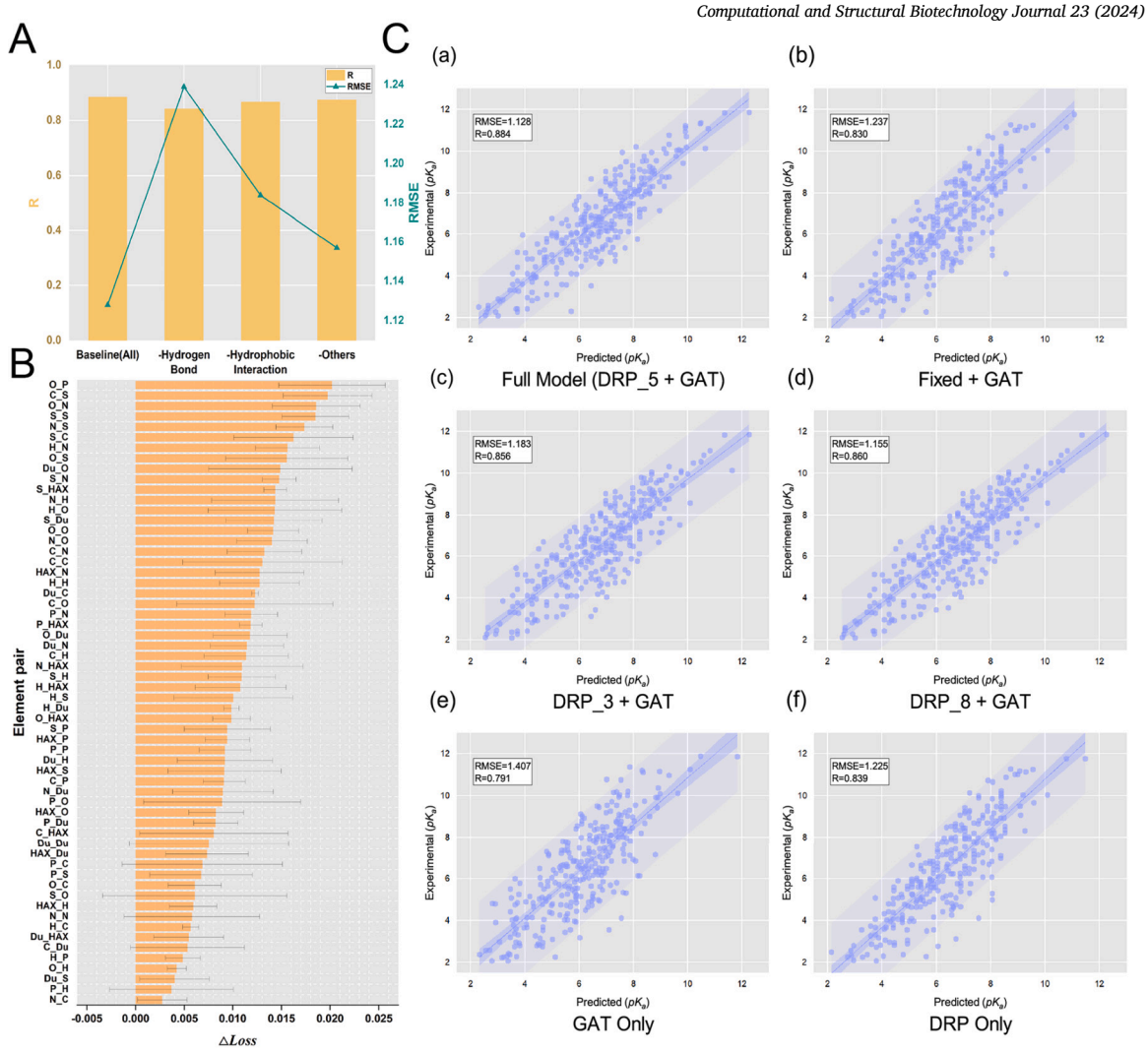

Fig. 3. Performance comparison of various models integrating dynamic radial partitioning (DRP) and graph attention network (GAT) in predicting pK_a. (A) RMSE and R-value of the model after removing different interaction types. (B) Contribution of different element pairs to model performance, represented by their impact on loss ($\Delta Loss$). (C) Scatter plots showing the predicted vs. experimental pK_a values for different model configurations: (a) Full Model (DRP_5 + GAT), (b) Fixed + GAT, (c) DRP_3 + GAT, (d) DRP_8 + GAT, (e) GAT Only, and (f) DRP Only.

4.3. Element-pair importance analysis

Fig. 3B illustrates the change in model loss relative to the optimal model upon the systematic removal of each elemental interaction pair. To investigate the impact of different combinations of element pairs on model performance, we sequentially eliminated one pair at a time and observed the resulting variations in model metrics. Overall, the removal of individual element pairs led to minor alterations in performance. However, the exclusion of specific element combinations—such as O_P, C_S, O_S, and H_N—that represent key biological interactions like hydrogen bonding and hydrophobic interactions resulted in a significant decrease in model performance compared to other combinations. The removal of other element pairs had a less pronounced effect, suggesting they may be involved in weaker or non-critical interactions. Furthermore, element combinations associated with larger errors in the analysis may be crucial to model performance in certain contexts, depending on factors such as the complexity of the molecular structure or the characteristics of specific binding sites.

4.4. Ablation studies

To validate the contribution of DRP to the model's performance, we conducted a systematic series of ablation experiments. These experiments were designed to analyze the impact of individual components on prediction outcomes by progressively removing or replacing different feature extraction modules. The experimental setup comprised the following configurations:

(1) Retention of both dynamic radial partitioning and molecular graph features: In this configuration, ligand–protein complexes were modeled using a dynamic layer feature extraction method combined with GAT features. The number of layers was dynamically adjusted by the K-means clustering algorithm based on actual interatomic distances. This setup served to verify the effectiveness of dynamic layer division, providing a foundation for subsequent performance evaluations and comparisons with other state-of-the-art methods.

(2) Use of fixed-distance partitioning with GAT features: We employed a fixed-distance division to extract interaction features and fused them with GAT features. Interatomic distances were segmented into multiple fixed intervals (e.g., 0–2Å, 2–4Å, 4–6Å) and combined with GAT features. This experiment evaluated the effectiveness of DRP with fixed-distance segmentation, offering insights into the advantages of adaptive feature extraction methods.

(3) Variation in the number of dynamic layers: Dynamic layer feature extraction and GAT feature fusion were tested with different numbers of layers. The model was configured with 3, 5, and 8 layers to assess performance under varying layer counts. This experiment aimed to optimize

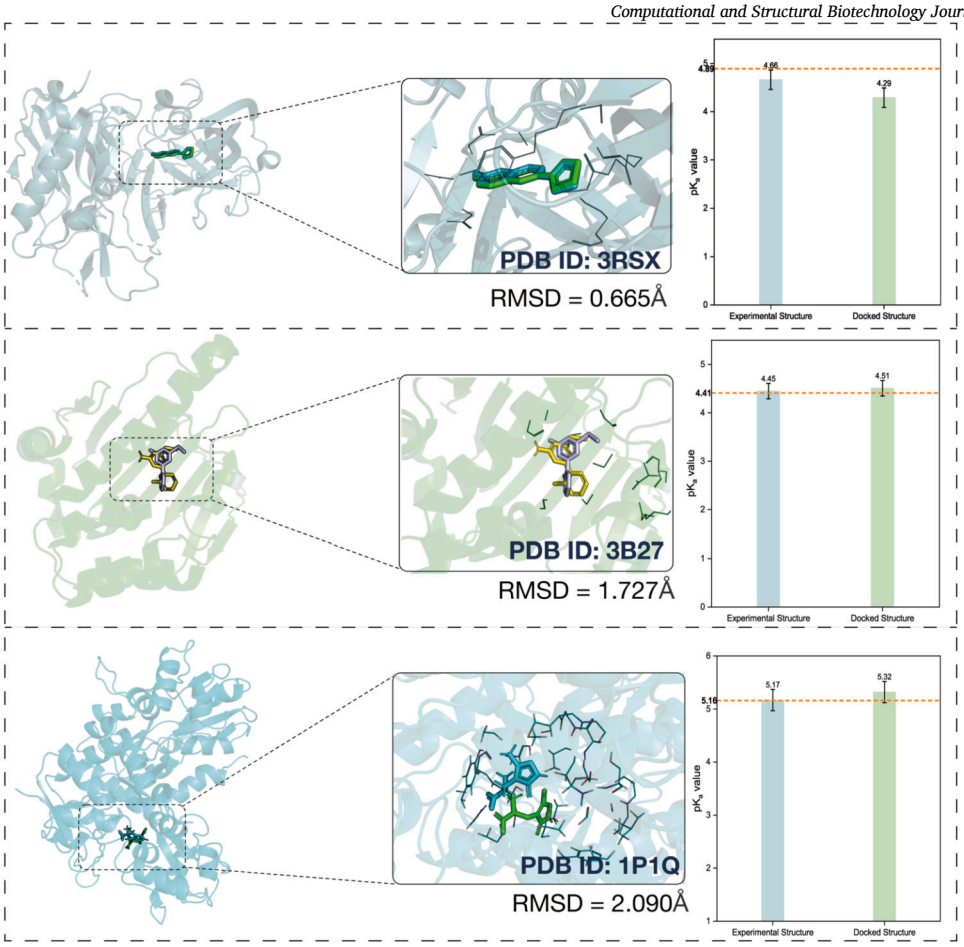

Fig. 4. Comparison of Experimental and Docked Structures: Structural Alignment and Predicted Binding Affinities.

Table 2
Performance Comparison of Different Model Configurations on CASF-2016 Core Set.

Model	DRP Layers	RMSE	R
DRP + GAT	5	1.128	0.884
Fixed + GAT	5	1.237	0.830
DRP + GAT	3	1.183	0.856
DRP + GAT	8	1.155	0.860
GAT Only	-	1.407	0.791
DRP Only	5	1.225	0.839

the selection of the most effective number of layers, ensuring an optimal balance between computational efficiency and predictive accuracy.

(4) Independent removal of DRP and GAT components: We separately removed the DRP and GAT modules to evaluate the individual contributions of dynamic interaction features and molecular structure features to the model's performance, as well as to assess their complementarity.

Table 2 and Fig.3C summarize the experimental results on the CASF-2016 core set. The findings demonstrate that both dynamic spatial distance features and molecular graph features are indispensable for our method.

Table 2 illustrates that the model integrating DRP feature extraction with GAT features (Case 1) achieved the highest performance in terms of both RMSE and R, significantly outperforming the traditional fixed DRP feature extraction method (Experiment Group 2). This improvement demonstrates that the DRP approach captures intricate protein-ligand interactions more effectively. Across multiple experiments, the optimal configuration was consistently found to be a five-layer model, with an RMSE of 0.87 and R of 0.91. This layer depth appears to balance the need for capturing interaction information without overfitting or introducing redundancy. Additionally, results from Experiment Groups 5 and 6 show that combining DRP and GAT features outperforms using either feature alone. While GAT features effectively capture molecular structural information, their integration with DRP features provides a more comprehensive representation of 3D protein–ligand interactions, thereby improving the model's predictive accuracy.

4.5. Robustness of MM-DRPNet

In this study, we assessed the robustness of our model by evaluating its performance on docking-generated poses derived from the CASF-2016 dataset. This dataset consists of experimentally determined protein–ligand complexes, which we re-docked using AutoDock Vina to generate multiple docking poses for each complex. Native-like conformations were selected based on their RMSD values, applying a threshold of 2Å from the crystal structures. These selected docking poses were then input into our model for binding affinity prediction. To evaluate the consistency and robustness of the model, we compared the predicted affinities from the docking poses with those obtained from the experimental structures (The correlation between predicted and experimental pK_a values for the docked structures is shown in Supporting Information Figure S3.). Fig. 4 presents a comparative analysis of structural alignments and pK_a values between experimental and docked structures for three protein–ligand complexes (PDB IDs: 3RSX, 3B27 and 1P1Q). On the left, the 3D representations display both experimental and docked ligand conformations within the protein binding pockets, with RMSD values indicating the structural differences between the two poses. On the right, bar charts illustrate the predicted pK_a values for each complex,

revealing minimal differences between the experimental and docked structures. These findings demonstrate the model's robustness in accommodating conformational variations while accurately predicting binding affinities.

5. Conclusion

In this study, we have introduced a novel multimodal framework, MM-DRPNet, for predicting protein–ligand binding affinity by integrating structural and interaction features from both proteins and ligands. The core innovation of our approach lies in the introduction of the Dynamic Radial Partitioning, which captures spatial interaction features within three-dimensional protein–ligand complexes. By systematically incorporating multiple interaction types—such as hydrogen bonds, hydrophobic interactions, and other physicochemical interactions—into our feature extraction process, we extracted highly informative features from the 3D structures. These features, combined with molecular graph representations of ligands, enable our model to precisely capture both global and local interaction patterns. The results demonstrate that MM-DRPNet significantly outperforms state-of-the-art deep learning models across multiple benchmark datasets, achieving superior predictive accuracy. The framework not only excels in performance but also exhibits robustness across varying interaction types, as evidenced by our ablation studies. Notably, the results highlight the critical role of hydrogen bonds and hydrophobic interactions in binding affinity prediction, aligning with established biochemical understanding.

The introduction of DRP as a feature extraction method offers a new avenue for exploring spatial relationships in protein–ligand complexes. This approach has the potential to be extended to other types of biomolecular interactions or integrated into different multimodal frameworks. While our model shows great promise, several directions for future research warrant exploration. First, validating MM-DRPNet's performance on emerging comprehensive datasets like PLINDER [52] would provide additional insights into its generalizability across diverse protein-ligand interactions. Second, comparing our approach with recent methods could reveal complementary strengths and potentially lead to more robust hybrid approaches. Additionally, exploring the model's applicability across a wider range of proteins and ligands, and investigating its integration into virtual screening pipelines remain important future directions. These extensions would further validate the model's utility in real-world drug discovery applications.

Overall, MM-DRPNet provides a powerful tool for computational drug discovery, with the potential to accelerate the identification and optimization of therapeutic compounds. By enhancing the accuracy of binding affinity predictions through our novel DRP approach and multimodal framework, our work contributes to more efficient drug development processes and advances the field of computational biology. Future validations and extensions of our model will further strengthen its practical impact in drug discovery applications.

CRediT authorship contribution statement

Dayan Liu: Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation, Conceptualization. **Tao Song:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Shudong Wang:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

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

Acknowledgements

This work was supported by National Key Research and Development Project of China (2021YFA1000102, 2021YFA1000103), Natural Science Foundation of China (Grant Nos. 61873280, 61972416), Taishan Scholarship (tsqn201812029), Foundation of Science and Technology Development of Jinan (201907116), Shandong Provincial Natural Science Foundation (ZR2021QF023), Fundamental Research Funds for the Central Universities (24CX04029A).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csbj.2024.11.050.

Data availability

The source code is available on https://github.com/Bigrock-dd/MMDRPv1. The data sets are also available: http://www.pdbbind.org.cn/download/CASF-2016.tar.gz.

References

[1] Jorgensen WL. The many roles of computation in drug discovery. Science 2004;303(5665):1813–8.
[2] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 2004;3(11):935–49.
[3] Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput-Aided Drug Des 2011;7(2):146–57.
[4] Huang S-Y, Zou X. Advances and challenges in protein-ligand docking. Int J Mol Sci 2010;11(8):3016–34.
[5] Mobley DL, Dill KA. Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". Structure 2009;17(4):489–98.
[6] Schneider G. Virtual screening: an endless staircase? Nat Rev Drug Discov 2010;9(4):273–6.
[7] Wu J, Chen H, Cheng M, Xiong H. Curvagn: curvature-based adaptive graph neural networks for predicting protein-ligand binding affinity. BMC Bioinform 2023;24(1):378.
[8] Wang W, Sun B, Liu D, Wang X, Zhang H. Dpla: prediction of protein-ligand binding affinity by integrating multi-level information. In: 2021 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE; 2021. p. 3428–34.
[9] Nguyen DD, Wei G-W. Agl-score: algebraic graph learning score for protein–ligand binding scoring, ranking, docking, and screening. J Chem Inf Model 2019;59(7):3291–304.
[10] Xu S, Shen L, Zhang M, Jiang C, Zhang X, Xu Y, et al. Surface-based multimodal protein-ligand binding affinity prediction. Bioinformatics 2024:btae413.
[11] Holland PL. Introduction: reactivity of nitrogen from the ground to the atmosphere. Chem Rev 2020;120(12):4919–20.
[12] Schindler CE, Baumann H, Blum A, Böse D, Buchstaller H-P, Burgdorf L, et al. Large-scale assessment of binding free energy calculations in active drug discovery projects. J Chem Inf Model 2020;60(11):5457–74.
[13] Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. Preprint. arXiv:1710.10903, 2017.
[14] Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput-Aided Mol Des 2013;27:221–34.
[15] Trott O, Olson AJ. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31(2):455–61.
[16] Ganesan A, Coote ML, Barakat K. Molecular dynamics-driven drug discovery: leaping forward with confidence. Drug Discov Today 2017;22(2):249–69.
[17] Cozzini P, Fornabaio M, Marabotti A, Abraham DJ, Kellogg GE, Mozzarelli A. Free energy of ligand binding to protein: evaluation of the contribution of water molecules by computational methods. Curr Med Chem 2004;11(23):3093–118.
[18] Krishnan SR, Bung N, Bulusu G, Roy A. Accelerating de novo drug design against novel proteins using deep learning. J Chem Inf Model 2021;61(2):621–30.
[19] Chen X, Yan C-C, Zhang X, Li Z-L, Deng L, Zhang Y. Drug–target interaction prediction: databases, web servers and computational models. Brief Bioinform 2016;17(4):696–712.
[20] Wen M, Zhang Z-H, Niu S, Sha H-Y, Yang R-F, Yun Y-H, et al. Deep-learning-based drug–target interaction prediction. J Proteome Res 2017;16(4):1401–9.
[21] Öztürk H, Özgür A, Ozkirimli E. Deepdta: deep drug–target binding affinity prediction. Bioinformatics 2018;34(17):i821–9.
[22] Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug–target interaction. Bioinformatics 2011;27(21):3036–43.

[23] Bleakley K, Yamanishi Y. Supervised prediction of drug–target interactions using bipartite local models. Bioinformatics 2009;25(18):2397–403.
[24] Marcou G, Rognan D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. J Chem Inf Model 2007;47(1):195–207.
[25] Deng Z, Chuaqui C, Singh J. Structural interaction fingerprint (sift): a novel method for analyzing three-dimensional protein- ligand binding interactions. J Med Chem 2004;47(2):337–44.
[26] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, 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):1739–49.
[27] Mamoshina P, Vieira A, Putin E, Zhavoronkov A. Applications of deep learning in biomedicine. Mol Pharm 2016;13(5):1445–54.
[28] Rao R, Liu J, Verkuil R, Meier J, Canny J, Abbeel P, et al. Msa transformer. bioRxiv. https://doi.org/10.1101/2021.02.12.430858, 2021.
[29] Elnaggar A, Heinzinger M, Dallago C, Rehawi G, Yu W, Jones L, et al. Prottrans: towards cracking the language of life's code through self-supervised deep learning and high performance computing. IEEE Trans Pattern Anal Mach Intell 2021. https://doi.org/10.1109/TPAMI.2021.3095381.
[30] Bepler T, Berger B. Learning protein sequence embeddings using information from structure. In: International conference on learning representations; 2019.
[31] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[32] Sønderby SK, Sønderby CK, Nielsen H, Winther O. Convolutional lstm networks for subcellular localization of proteins. In: 2015 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE; 2015. p. 1397–402.
[33] Alley EC, Khimulya G, Biswas S, AlQuraishi M, Church GM. Unified rational protein engineering with sequence-based deep representation learning. Nat Methods 2019;16(12):1315–22. https://doi.org/10.1038/s41592-019-0598-1.
[34] Strodthoff N, Wagner P, Wenzel M, Samek W. Deep learning in ecg analysis: benchmarks and insights from ptb-xl. IEEE J Biomed Health Inform 2020;25(5):1519–28. https://doi.org/10.1109/JBHI.2020.3022989.
[35] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Advances in neural information processing systems; 2017. p. 5998–6008.
[36] Rives A, Meier J, Sercu T, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci 2021;118(15):e2016239118.
[37] Xu Y, Wang S, Hu J, Xue B, Cao R, Wang Z, et al. Deep learning for protein fold recognition: an overview. IEEE Access 2020;8:33650–66. https://doi.org/10.1109/ACCESS.2020.2974204.
[38] Koh HY, Nguyen AT, Pan S, May LT, Webb GI. Psichic: physicochemical graph neural network for learning protein-ligand interaction fingerprints from sequence data. 2023. bioRxiv.
[39] Nguyen TD, Le HH, Quinn TP, Nguyen TH. Graphdta: predicting drug–target binding affinity with graph neural networks. Bioinformatics 2021;37(8):1140–7.
[40] Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang S, et al. Transformercpi: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label smoothing. Bioinformatics 2020;36(16):4406–14.
[41] Zheng S, Li Y, Chen S, Xu J, Yang Y. Onionnet: a multiple-layer intermolecular-contact-based convolutional neural network for protein–ligand binding affinity prediction and binding pose classification. J Chem Inf Model 2019;59(10):4381–8.
[42] Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P. Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. Bioinformatics 2018;34(21):3666–74.
[43] Jiménez J, Skalic M, Martinez-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):287–96.
[44] Tran-Nguyen V-K, Bret G, Rognan D. True accuracy of fast scoring functions to predict high-throughput screening data from docking poses: the simpler the better. J Chem Inf Model 2021;61(6):2788–97.
[45] Ragoza M, Hochuli J, Idrobo E, Sunseri J, Koes DR. Protein–ligand scoring with convolutional neural networks. J Chem Inf Model 2017;57(4):942–57.
[46] Wallach I, Dzamba M, Heifets A. Atomnet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. Preprint. arXiv:1510.02855, 2015.
[47] Wang R, Fang X, Lu Y, Wang S. The pdbbind database: collection of binding affinities for protein- ligand complexes with known three-dimensional structures. J Med Chem 2004;47(12):2977–80.
[48] Gaillard T. Evaluation of autodock and autodock vina on the casf-2013 benchmark. J Chem Inf Model 2018;58(8):1697–706.
[49] Su M, Yang Q, Du Y, Feng G, Liu Z, Li Y, et al. Comparative assessment of scoring functions: the casf-2016 update. J Chem Inf Model 2018;59(2):895–913.
[50] Wang X, Pan Y. Deep learning models for protein–ligand binding affinity prediction. Curr Opin Struct Biol 2021;67:170–7. https://doi.org/10.1016/j.sbi.2021.01.009.
[51] Macari G, Toti D, Pasquadibisceglie A, Polticelli F. Dockingapp rf: a state-of-the-art novel scoring function for molecular docking in a user-friendly interface to autodock vina. Int J Mol Sci 2020;21(24):9548.
[52] Durairaj J, Adeshina Y, Cao Z, Zhang X, Oleinikovas V, Duignan T, et al. Plinder: The protein-ligand interactions dataset and evaluation resource. 2024. bioRxiv.