



Cite this: DOI: 10.1039/d5dd00365b

# MolRes-DTA: a molecular-multiview fusion and residue-aware model for drug-target affinity prediction

Hongli Hou, <sup>ab</sup> Qi Wei, <sup>b</sup> Dian Huang, <sup>\*b</sup> Minglu Zhao, <sup>b</sup> Hongliang Duan <sup>a</sup>  
and Shengzhong Feng <sup>\*b</sup>

Accurate prediction of drug-target affinity (DTA) is crucial for drug screening and reducing drug development costs. Despite the significant progress made by deep learning methods in DTA prediction, most existing approaches neglect two critical factors: the influence of drug molecular size and the different contribution of amino acid residues to DTA. Here, we propose an affinity prediction model, MolRes-DTA, which introduces multiview drug characterization and a dynamic residue-aware network to capture the influence of molecular size on affinity prediction and weigh the contributions of different residues. Experiments on the Davis and KIBA datasets demonstrate that MolRes-DTA outperforms baseline models by 15.58% and 20.11%, respectively. Further analysis shows that our multiview drug representation improves prediction accuracy across different types of molecular sizes, with particularly notable gains for larger compounds. To our knowledge, this is the first study to explore the impact of molecular size on DTA prediction, providing a novel perspective for enhancing the accuracy of DTA prediction.

Received 16th August 2025

Accepted 8th April 2026

DOI: 10.1039/d5dd00365b

rsc.li/digitaldiscovery

## 1 Introduction

Accurate prediction of drug-target binding affinity (DTA) is essential for both drug discovery and drug repurposing.<sup>1–6</sup> Although experimental assays generate reliable affinity measurements, they are inherently time-consuming, costly, and difficult to scale. Classical computational approaches such as molecular docking and molecular dynamics simulations can offer mechanistic insight into ligand–receptor interactions,<sup>7,8</sup> but they rely on high-quality structural data and are often impractical for large-scale screening.<sup>9</sup> These have motivated the development of more efficient computational methods that enable high-throughput evaluation of extensive compound libraries.<sup>10–13</sup>

The increasing availability of DTA datasets and rapid progress in machine learning have expanded the scope and capability of affinity prediction.<sup>14</sup> Early machine learning studies relied on handcrafted descriptors for both drugs and proteins, whereas deep learning has introduced more expressive feature extractors and more flexible nonlinear interaction models. Recent work has advanced along two main directions: the expansion of input modalities, which broadens from linear sequences to graphs structural representations, and the

innovation of model architectures, which captures the increasingly complex biochemical relationships.<sup>15,16</sup>

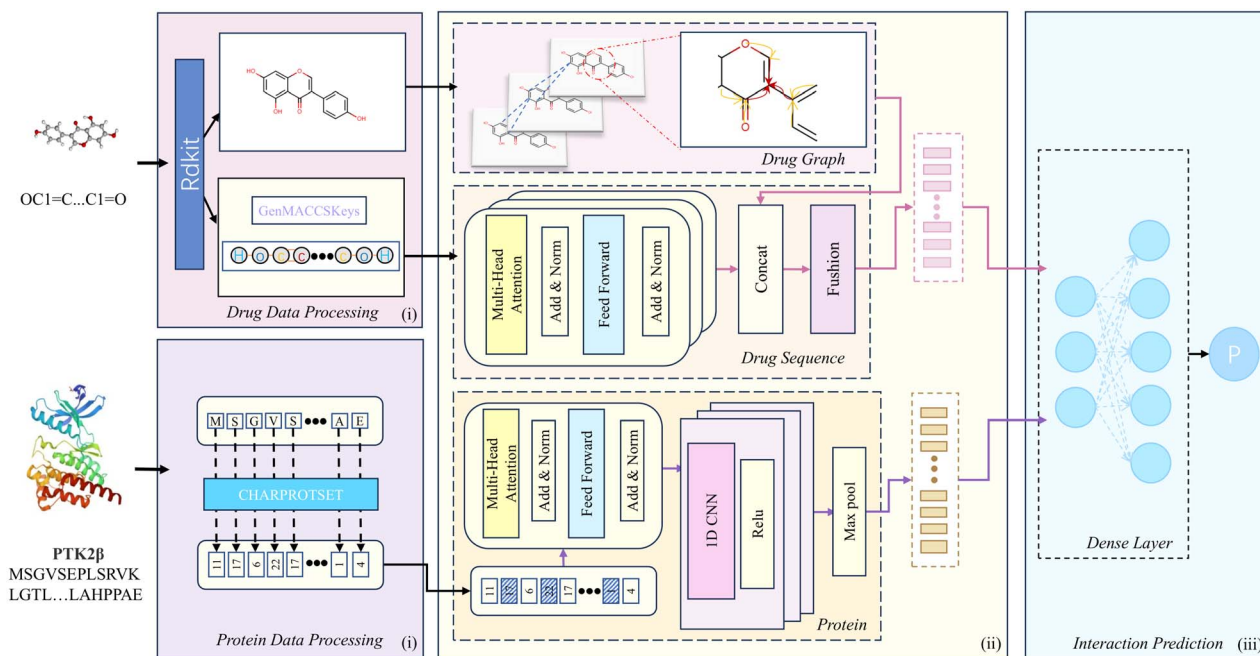
From a feature perspective, DTA prediction has progressively expanded information dimensions with the rapid advancement of deep learning methods.<sup>17</sup> Early SMILES-based works, DeepDTA and WideDTA, treated drugs as one-dimensional sequences,<sup>18,19</sup> and relied on convolutional or language-inspired encoders to capture local sequence motifs. Subsequent works such as ArkDTA, FingerDTA and TEFDTA, introduced molecular fingerprints to inject domain knowledge and improve chemical interpretability, demonstrating how pre-defined substructure encodings and global descriptors provide complementary information to sequence-derived features.<sup>20–22</sup>

As a further advance in DTA prediction, the adoption of explicit graph representations of small molecules, exemplified by GraphDTA and NG-DTA, has shown that graph neural networks can capture atomic topology and bonding patterns more effectively than one-dimensional encodings.<sup>23,24</sup> More recent graph-centered methods emphasize richer intramolecular relations and multi-scale interactions. For example, TDGraphDTA and MDCT-DTA incorporated modules for substructure interaction and enhanced node-level expressivity.<sup>25,26</sup> To jointly model chemical and structural determinants of binding, PocketDTA extend this line of work by combining atomic-level graph connectivity with three-dimensional pocket-aware information.<sup>27</sup> Although these developments, they collectively reveal a central challenge: how to capture atomic-level topological connectivity while also

<sup>a</sup>Faculty of Applied Sciences, Macao Polytechnic University, R. de Luís Gonzaga Gomes, 999078, Macao, China

<sup>b</sup>Guangdong Institute of Intelligence Science and Technology, Hengqin, Zhuhai, 519031, Guangdong, China. E-mail: huangdian@gdiist.cn; fengshengzhong@gdiist.cn





**Fig. 1** An overview of the MolRes-DTA architecture. (i) Multiview data encoding: drug SMILES strings are converted into molecular graphs and MACCS fingerprints, while protein sequences are encoded at the character level. (ii) Feature extraction module: a hybrid graph neural network and transformer are used to fuse structural and sequential information for drugs. For proteins, DRN integrates self-attention and convolutional operations to enhance semantic and structural representations. (iii) Affinity prediction module: the final embeddings of drugs and proteins are concatenated and fed into a multilayer perceptron (MLP) to predict binding affinity scores.

acquiring global functional substructure information from complementary molecular modalities, and how to integrate these views to overcome the limitations of single-modal representations in functional and structural characterization. We further note that drug molecular size may affect pharmacokinetics and delivery, including administration route, metabolic clearance and tissue penetration,<sup>28</sup> and that such effects could in turn influence the behavior of DTA prediction models.

From a model perspective, protein feature modeling has proceeded remarkably from local capture to global optimization. Early convolutional encoders, such as MT-DTI and MATT-DTI, used multi-layer and three-layer CNNs respectively, to extract local residue motifs from primary sequences.<sup>29,30</sup> Later, to link local residue signals with overall sequence context, GDilatedDTA applied bidirectional LSTM to perform recurrent information interaction.<sup>31</sup> As the emergence of the Transformer method, TransformerCPI enabled explicit modeling of long-range dependencies through self-attention and in practice has improved the ability to localize interaction regions between protein sequence and ligand atoms.<sup>32</sup> Subsequently, to capture both short-range substructure features and long-range contextual relationships, hybrid architectures such as MT-DTA and RRGDTA combine convolutional blocks with attention mechanism for local-global feature extraction.<sup>33,34</sup> With the advancement of Large Language Models in recent years, AttentionMGT-DTA and LLMDTA used pretraining models ESM-2 to further enrich the protein embeddings and improve the generalization of downstream tasks.<sup>35,36</sup> Despite these advances, most existing methods treat amino acid residues as uniformly contributing

units, which can dilute signals from functionally critical residues.

To address the challenges identified above, we propose MolRes-DTA, a molecular-modality fusion and residue-aware model for DTA prediction. On the drug side, MolRes-DTA integrates a hybrid graph neural module, which captures fine-grained atomic topology with a fingerprint branch that encodes global functional semantics. This multiview fusion enables the model to represent both local bonding patterns and overall substructure chemistry, improving robustness for small and structurally complex molecules. On the protein side, we introduce a Dynamic Residue-aware Network (DRN) that combines learnable residue-wise modulation with self-attention and multi-scale convolution, allowing the model to amplify functionally relevant residues while attenuating less informative regions. An overview of the MolRes-DTA architecture is presented in Fig. 1.

Our contributions are summarized as follows:

- We present MolRes-DTA, a model that integrates a dual-modality drug encoder and a residue-aware protein network. On the drug side, the model combines atomic-level graph representations and fingerprint-derived global semantics to mitigate the limitations of single-view encodings. On the protein side, the Dynamic Residue-aware Network assigns context-dependent weights to residues to better highlight functionally important regions.

- We identify and characterize a relationship between drug SMILES length and DTA prediction performance. A molecular-size stratification analysis reveals performance gaps across



molecular classes and offers suggestions for designing models and datasets that are robust to molecular-size variation.

- We provide extensive experimental and biological validation, containing ablation studies, molecular-size analysis, visualization of DRN-derived residue importance, comparison with docking results, and molecular dynamics simulations. Experiment results show MolRes-DTA improves predictive accuracy while providing reasonable interpretability.

## 2 Materials and methods

### 2.1 Datasets

In this study, we used three publicly available benchmark datasets, Davis,<sup>37</sup> KIBA,<sup>38</sup> and BindingDB<sup>39</sup> to comprehensively evaluate the performance of MolRes-DTA. These datasets span diverse classes of protein targets and drug molecules, providing representative and challenging scenarios for affinity prediction.

The Davis dataset contains 30 056 drug-target interaction pairs involving 68 compounds and 442 proteins, with binding affinities measured by the dissociation constant ( $K_d$ ). We adopted the corrected protein sequences provided by Li *et al.*<sup>22</sup> to ensure data consistency and integrity.

The KIBA dataset includes 118 254 drug-target interaction pairs between 2111 small molecules and 229 protein targets. Binding affinities are represented by KIBA scores, which integrate multiple biochemical measurements, including  $K_d$ , the inhibition constant  $K_i$ , and the half-maximal inhibitory concentration  $IC_{50}$ . We used the preprocessed and categorized version published by Li *et al.*,<sup>22</sup> enabling consistent partitioning and fair model comparison.

The BindingDB dataset includes 2.7 million affinity records involving over 9000 proteins and 1.2 million small molecules. After removing ambiguous and duplicate entries by Li *et al.*,<sup>22</sup> the curated version used in this study comprises 80 324 compounds, 5561 targets, and 1 254 402 interactions.

To ensure a fair and reproducible comparison with the baseline, we adopted the dataset partitioning procedures from TEFDTA. For Davis and KIBA, we used the fixed train/test split lists released by TEFDTA, corresponding to a random interaction-pair split with an 5:1 train:test ratio. For BindingDB, we used the predefined split provided by the TEFDTA benchmark without re-splitting to maintain strict comparability. More detailed information is shown in Table S1.

### 2.2 Overview of MolRes-DTA

MolRes-DTA is a deep learning model that integrates multiview drug encoding with residue-aware protein modeling to predict binding affinity. Drugs are represented through molecular graphs and MACCS fingerprints to capture complementary structural and semantic features, the complementarity of which is visually demonstrated in Fig. S1. Protein sequences are encoded and refined using methods such as DRN, which models residue-aware contributions. The extracted features are fused and passed through a prediction module to generate an affinity score.

### 2.3 Drug feature extraction

**2.3.1 MACCS fingerprint representation.** To capture key structural features of drug molecules, we adopt the standardized MACCS structural fingerprinting method, which maps SMILES-encoded molecular structures into fixed-length binary vectors. This fingerprint comprises 166 bits, each indicating the presence or absence of a specific chemical substructure, and serves to characterize the topological and functional properties of the molecular scaffold:

$$MS(i) = \begin{cases} 1, & \text{if the } i\text{-th substructure is present} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

This representation provides fixed dimensionality and strong interpretability, effectively mitigating the interference caused by varying molecular lengths during model training.

To further capture dependencies between substructures, we introduce a Transformer encoder for global modeling. Specifically, the input fingerprint matrix is projected into query, key, and value spaces:

$$Q_D = MS \times W_Q, K_D = MS \times W_K, V_D = MS \times W_V \quad (2)$$

The multi-head self-attention mechanism is then computed as:

$$\text{Attention}(Q_D, K_D, V_D) = \text{Softmax}\left(\frac{Q_D K_D^T}{\sqrt{d_k}}\right) V_D \quad (3)$$

where  $W_Q$ ,  $W_K$ ,  $W_V$  are learnable weight matrices and  $d_k$  is the dimensionality of the key vectors.

Finally, a max-pooling operation is applied to the attention output to generate the fingerprint-level drug representation:

$$D_1 = \text{MaxPooling}(\text{Attention}(Q_D, K_D, V_D)) \quad (4)$$

This vector encodes contextual dependencies among molecular substructures and enhances the capacity of the model to represent functional fragments of the drug.

**2.3.2 Molecular graph representation.** To capture the topological structure of drug molecules, we first parse the SMILES strings into molecular graphs. Specifically, we use RDKit to extract five atomic attributes—atomic number, degree, formal charge, number of radical electrons, and aromaticity—as node features  $V$ , and construct an adjacency matrix  $E$  based on covalent bonds, forming an undirected graph  $G = (V, E)$ .

A GCN is employed to extract local structural features. At each GCN layer, the representation of a node is updated by aggregating information from its neighbors:

$$X'_i = F_{\text{GCN}}(G) \quad (5)$$

where  $X'_i$  denotes the updated embedding of node  $i$ , and  $X'_j$  represents the embeddings of its neighboring nodes. The propagation rule of GCN is based on the graph Laplacian and can be expressed as:

$$H^{(l+1)} = \sigma(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^{(l)} W^{(l)}) \quad (6)$$



where  $\tilde{A} = A + I$  is the adjacency matrix with self-loops,  $\tilde{D}$  is the corresponding degree matrix,  $H^{(l)}$  is the node representation at the  $l$ -th layer,  $W^{(l)}$  is a learnable weight matrix, and  $\sigma$  is the activation function.

After obtaining the initial node embeddings  $X_i'$  via GCN, we further apply a Graph Attention Network (GAT) to capture more complex, asymmetric dependencies between nodes. The GAT layer assigns different attention weights to neighbor nodes and is formulated as:

$$X_i'' = F_{\text{GAT}}\left(X_i', \{X_j'\}\right) \quad (7)$$

Specifically, for each neighboring node  $j$ , the attention coefficient is computed as:

$$e_{ij} = \text{LeakyReLU}\left(a^T [WX_i' \parallel WX_j']\right) \quad (8)$$

where  $W$  is the shared linear transformation matrix,  $a$  is the attention vector, and  $\parallel$  denotes vector concatenation. The attention scores are normalized using the softmax function:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})} \quad (9)$$

The final node representation is obtained as a weighted sum over its neighbors:

$$X_i'' = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} WX_j'\right) \quad (10)$$

After joint GCN and GAT encoding, we apply mean pooling over all node embeddings to obtain a graph-level representation  $D_2$ , which summarizes atomic-level information into a global structural embedding.

To construct the final drug representation, we concatenate the fingerprint-based feature vector  $D_1$  with the graph-level embedding  $D_2$ , and apply a fully connected transformation with a non-linear activation:

$$D = \text{ReLU}(W_{\text{fusion}}[D_1 \parallel D_2] + b_{\text{fusion}}) \quad (11)$$

where  $W_{\text{fusion}}$  and  $b_{\text{fusion}}$  denote the trainable weights and bias of the fusion layer. This fused representation effectively integrates both molecular substructure fingerprints and topological graph information, and is used as input for subsequent affinity prediction tasks.

**2.3.3 Protein feature extraction.** To process the input protein sequence, we first map each amino acid (*e.g.* A, C, E, *etc.*) to a unique integer using a predefined vocabulary. This integer sequence is then transformed into a learnable embedding matrix  $p$ , where each row corresponds to the embedding vector of an amino acid. All sequences are truncated or padded to a fixed length of 1000 to ensure uniform input dimensions.

Traditional protein encoding methods often assume uniform contributions from all residues, ignoring the functional diversity among amino acids. To address this, we introduce DRN that applies a learnable modulation matrix  $G$  to assign distinct weights to each residue across its feature dimensions, where  $g \in [\gamma_{\min}, 1.0]$ . In practice,  $G$  is initialized as a channel-wise learnable mask that is jointly optimized with the network parameters, enabling the model to adaptively scale

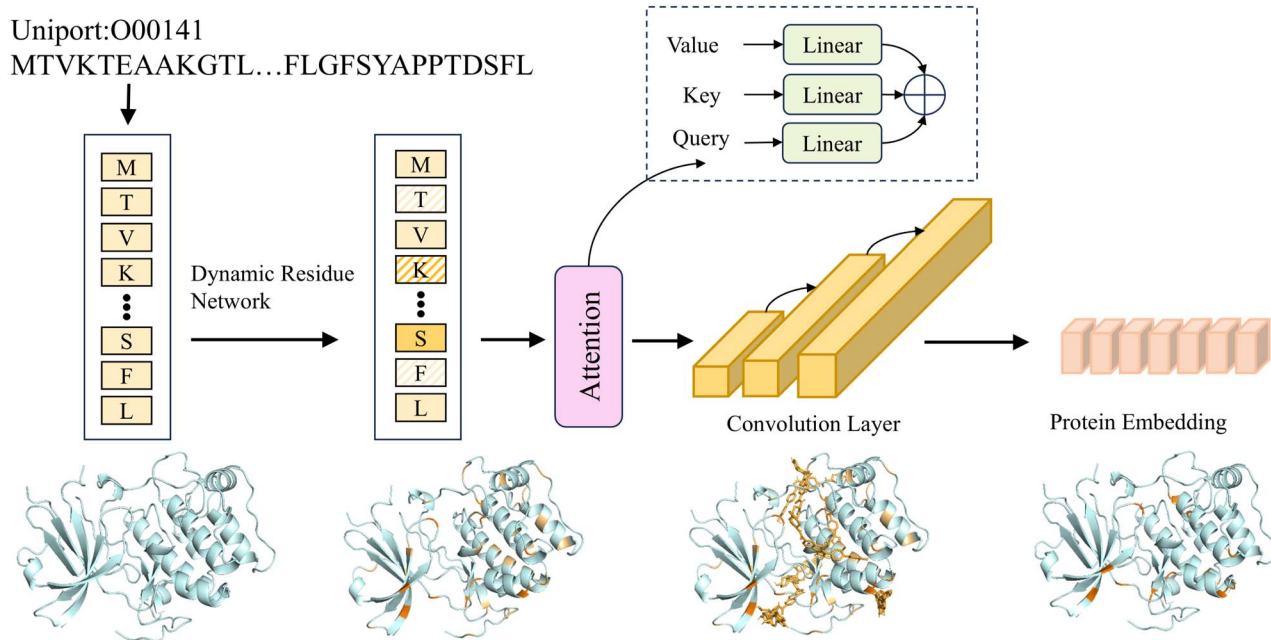


Fig. 2 Visualization of the protein feature extraction process and residue-aware modeling. The upper panel illustrates the overall protein encoding pipeline, including initial embedding, DRN, self-attention refinement, and multi-scale convolution. The lower panel visualizes the evolution of residue-level importance across layers. Color intensity reflects the relative contribution of each amino acid residue to the learned representation, where darker shades indicate residues receiving higher attention due to their contextual or structural relevance.



each residue's embedding according to its contextual relevance. This lightweight parameterization ensures fine-grained weighting without adding architectural complexity. The reweighted feature is computed as:

$$p' = G \odot p \quad (12)$$

where  $\odot$  denotes multiplication by element. Residues assigned higher weights retain more representational information, whereas those with lower weights are suppressed. To illustrate how the model learns to emphasize informative residues throughout the encoding process, we visualize the evolution of the residue weights in layers, highlighting the influence of the DRN module (Fig. 2).

The weighted embeddings  $p'$  are passed to a self-attention module to capture global dependencies among residues:

$$P_T = \text{Softmax} \left( \frac{p' W_Q (p' W_K)^\top}{\sqrt{d_k}} \right) p' W_V \quad (13)$$

where  $W_Q$ ,  $W_K$ ,  $W_V$  are trainable projection matrices, and  $d_k$  is the dimension of the key vectors. This mechanism allows the network to focus on semantically meaningful regions of the sequence, thereby enhancing representation learning.

To further capture local structural interactions, we apply three layers of 1D convolution after attention:

$$C_0 = P_T, C_i = \text{ReLU}(\text{Conv1D}(C_{i-1}, W_i)), i = 1, 2, 3 \quad (14)$$

where  $W_1$ ,  $W_2$ ,  $W_3$  are convolutional kernels with different receptive fields. This convolutional module captures local contextual interactions between adjacent residues, serving as a complementary enhancement to the attention mechanism.

Finally, global max pooling is applied to the output of the final convolutional layer to capture the most salient activations across channels, yielding the protein-level embedding  $P$ . This representation effectively integrates both global sequence dependencies and local structural features, and serves as input for downstream DTA prediction tasks.

**2.3.4 Affinity prediction.** The final drug representation  $D$  and protein representation  $P$  are concatenated to form a unified input vector  $v = [D||P]$ , which is then fed into a MLP to model complex non-linear interactions and predict binding affinity. The hidden representation at the  $l$ -th layer is computed as:

$$\hat{h}^{(l)} = \begin{cases} \sigma(W^{(l)} \hat{h}^{(l-1)} + b^{(l)}), 1 \leq l < L \\ W^{(L)} \hat{h}^{(L-1)} + b^{(L)}, l = L \end{cases}$$

with  $\hat{h}^{(0)} = v$  (15)

Here,  $L$  denotes the total number of layers,  $\sigma$  is the LeakyReLU activation function used in hidden layers, and the final layer performs linear regression to yield the predicted affinity score:

$$\hat{y} = \hat{h}^{(L)} \quad (16)$$

MolRes-DTA is optimized using mean squared errors (MSE) loss function:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (17)$$

where  $y_i$  and  $\hat{y}_i$  denote the ground truth and predicted affinity values for the  $i$ -th sample, respectively, and  $N$  is the total number of samples. This loss function effectively measures the discrepancy between predicted and actual values and guides the parameter updates through backpropagation to minimize the overall prediction error.

**2.3.5 Evaluation metrics.** The performance of the proposed model was evaluated using three metrics: MSE, CI, and  $r_m^2$ .

MSE is a widely used metric in regression tasks that measures the squared difference between predicted values and the truth of the ground. It is defined as:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (Y_i - Y'_i)^2 \quad (18)$$

where  $Y_i$  is the ground truth value for the  $i$ -th sample,  $Y'_i$  is the corresponding predicted value, and  $N$  is the total number of samples. A lower MSE indicates a better model fit. However, this metric is sensitive to outliers and thus should be interpreted in conjunction with other evaluation metrics.

Concordance Index (CI) assesses the consistency between the predicted and actual value ranking. Evaluates the model's ability to correctly rank sample pairs and is defined as:

$$\text{CI} = \frac{1}{Z} \sum_{\delta_j > \delta_i} h(Y'_i - Y'_j) \quad (19)$$

where  $Z$  is the number of comparable sample pairs satisfying  $\delta_j > \delta_i$ , and  $h(x)$  is an indicator function defined as:

$$h(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0.5 & \text{if } x = 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (20)$$

CI ranges from 0 to 1, with higher values indicating stronger ranking performance of the model.

$r_m^2$  is a metric designed to evaluate the degree of regression toward the mean and to assess the external predictive ability of the model. It helps determine whether the model suffers from overfitting or underfitting. It is computed as follows:

$$r_m^2 = r^2 \times \left( 1 - \sqrt{r^2 - r_0^2} \right) \quad (21)$$

where  $r^2$  is the coefficient of determination based on least-squares regression, and  $r_0^2$  is the squared correlation coefficient from regression constrained to pass through the origin. A model is typically considered acceptable when  $r_m^2 \geq 0.5$ . This metric effectively reflects the model's generalization ability of and serves as a reference for model architecture and parameter selection.

**2.3.6 Implementation details.** The experiments were conducted on a system equipped with an Intel(R) Xeon(R) Silver



Table 1 Hyperparameter settings used in the experiments

Hyperparameter	Setting
Learning rate	0.0001
Dropout rate	0.1
GCN layers	3
MLP layers	2
Transformer layers	3
Attention heads	8
Optimizer	Adam
Batch size	256
Epochs	300

4210R CPU @ 2.40 GHz, dual Tesla V100 GPUs (32 GB VRAM each) and running Ubuntu 22.04.5. The model was implemented in Python 3.9 with CUDA 12.4 support, using PyTorch 2.5.1 as the deep learning model. The hyperparameter settings used for the training are summarized in Table 1.

## 3 Results and discussion

### 3.1 Performance comparison with baseline models

We compare our model with several representative methods, relying on results reported in their original publications to eliminate potential discrepancies arising from reimplementations. As summarized in Tables 2 and 3, MolRes-DTA achieves competitive performance across all evaluation metrics on both the Davis and KIBA datasets, with our results presented as the mean  $\pm$  standard deviation over five independent random seeds. Specifically, on the Davis dataset, the model reduces the MSE to 0.167, presenting a 15.58% improvement over the baseline TEFDTA (MSE = 0.199). In addition, it achieves CI and  $r_m^2$  scores of 0.908 and 0.784, respectively. These results demonstrate that MolRes-DTA performs well in the Davis dataset where the affinity distribution is more densely populated. We also conducted an additional experiment on the Davis dataset using an alternative splitting approach, with results presented in Table S2.

In contrast, the KIBA dataset presents greater challenges due to its broader distribution of binding affinities. Even under

Table 2 Performance comparison on the Davis dataset (best results in bold)

Method	MSE ( $\downarrow$ )	CI ( $\uparrow$ )	$r_m^2$ ( $\uparrow$ )
KronRLS <sup>40</sup>	0.379	0.781	0.407
SimBoost <sup>41</sup>	0.282	0.872	0.644
DeepDTA <sup>18</sup>	0.261	0.878	0.630
DeepCDA <sup>42</sup>	0.248	0.891	0.649
MATT-DTI <sup>30</sup>	0.229	0.890	0.682
GraphDTA <sup>23</sup>	0.233	0.890	0.747
GDilatedDTA <sup>31</sup>	0.232	0.885	0.686
TEFDTA <sup>22</sup>	0.199	0.890	0.756
PocketDTA <sup>27</sup>	0.177 $\pm$ 0.013	0.903 $\pm$ 0.005	0.731 $\pm$ 0.017
AttentionMGT-DTA <sup>35</sup>	0.193 $\pm$ 0.001	0.891 $\pm$ 0.005	0.699 $\pm$ 0.027
LLMDTA <sup>36</sup>	0.226 $\pm$ 0.001	0.884 $\pm$ 0.001	0.717 $\pm$ 0.010
MolRes-DTA (ours)	<b>0.167 <math>\pm</math> 0.001</b>	<b>0.908 <math>\pm</math> 0.001</b>	<b>0.784 <math>\pm</math> 0.004</b>

Table 3 Performance comparison on the KIBA dataset (best results in bold)

Method	MSE ( $\downarrow$ )	CI ( $\uparrow$ )	$r_m^2$ ( $\uparrow$ )
KronRLS	0.411	0.782	0.342
SimBoost	0.222	0.836	0.629
DeepDTA	0.194	0.863	0.673
DeepCDA	0.176	0.890	0.682
MATT-DTI	0.151	0.889	0.756
GraphDTA	0.158	0.887	0.674
GDilatedDTA	0.156	0.876	0.775
TEFDTA	0.184	0.860	0.731
PocketDTA	0.140 $\pm$ 0.004	0.892 $\pm$ 0.002	0.771 $\pm$ 0.011
AttentionMGT-DTA	<b>0.140 <math>\pm</math> 0.001</b>	<b>0.893 <math>\pm</math> 0.001</b>	<b>0.786 <math>\pm</math> 0.018</b>
LLMDTA	0.162 $\pm$ 0.001	0.872 $\pm$ 0.001	0.768 $\pm$ 0.001
MolRes-DTA (ours)	0.147 $\pm$ 0.001	<b>0.893 <math>\pm</math> 0.001</b>	0.778 $\pm$ 0.006

these more complex conditions, MolRes-DTA attains an MSE of 0.147, outperforming TEFDTA (0.184) and GDilatedDTA (0.156) with relative reductions of 20.1% and 5.8%. While PocketDTA and AttentionMGT-DTA obtain slightly lower MSE values, MolRes-DTA attains a tied-best CI of 0.893 and a competitive  $r_m^2$  of 0.778, indicating that the predictive ranking and reliability of the model remain state-of-the-art. Combined with the leading results on the Davis dataset, these findings highlight the robustness and general applicability of MolRes-DTA across affinity datasets of varying complexity. The model's consistent superiority across datasets with varying levels of complexity underscores its robustness in DTA prediction. Additional significant analysis are presented in Table S3.

To further validate the prediction accuracy of MolRes-DTA, we analyzed the distribution of predicted and actual values. In Fig. 3(a), the Davis dataset results indicate that most data points are highly concentrated around the reference line, especially in the affinity range of [5, 7], where the predictions closely match the true values. This suggests that MolRes-DTA exhibits a strong fitting capability in this interval and accurately captures the interactions of common binding strength samples. In the low-affinity region (<6), the model still demonstrates relatively stable predictions, although some deviations are observed. This may be partly attributed to the fact that affinity values below 5 were rounded to 5 during the dataset construction process, thereby reducing the prediction accuracy of this region.<sup>41</sup>

Fig. 3(b) presents the results on the KIBA dataset, where the predicted values exhibit a similarly strong linear trend and are densely distributed within the range [10, 14], after filtering out individual samples with very low affinity to improve visualization. Compared to the Davis dataset, fewer outliers are observed in KIBA, highlighting the robustness and generalization capability of MolRes-DTA under larger and more complex distribution scenarios. Overall, the compact scatter distribution and consistent trends across both datasets confirm the high predictive precision and stability of the model in binding affinity prediction tasks.

We further evaluate the generalization ability of MolRes-DTA on the BindingDB dataset, which contains a wider range of drug-target combinations and more diverse molecular



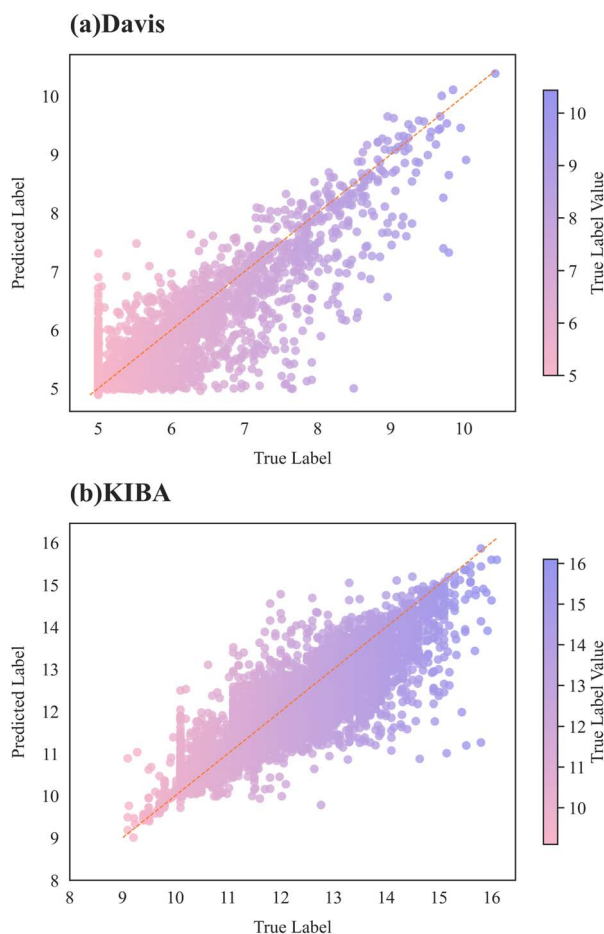


Fig. 3 Scatter plot on the (a) Davis and (b) KIBA dataset.

Table 4 Performance comparison on the BindingDB dataset

Method	MSE ( $\downarrow$ )	CI ( $\uparrow$ )	$r_m^2$ ( $\uparrow$ )
DeepDTA	0.812	0.795	0.618
DeepCDA	0.832	0.811	0.628
TEFDA	0.701	0.814	0.631
MolRes-DTA	<b>0.573 <math>\pm</math> 0.002</b>	<b>0.837 <math>\pm</math> 0.001</b>	<b>0.727 <math>\pm</math> 0.002</b>

structures than the previous two benchmark datasets. Following the experimental settings reported in TEFDTA for consistency, we ensured a fair comparison. As summarized in Table 4, MolRes-DTA consistently outperforms all baselines across evaluation metrics, achieving an 18.1% reduction in MSE relative to TEFDTA, along with notable improvements in CI and  $r_m^2$ . These results confirm that the model maintains strong performance under more complex structural variations and data distributions, supporting the overall effectiveness of its multiview fusion and residue-aware representation strategies.

### 3.2 Ablation study

We conducted comparisons of several variants of our proposed MolRes-DTA model on the Davis dataset to assess the contribution of core components: only MACCS, which uses molecular

Table 5 Ablation study of different modules on the Davis dataset

Method	MSE ( $\downarrow$ )	CI ( $\uparrow$ )	$r_m^2$ ( $\uparrow$ )
Only MACCS	0.228	0.882	0.730
Only graph	0.221	0.890	0.747
MACCS + graph	0.190	0.902	0.763
+Residue-aware	0.178	0.906	0.772
+Attention	<b>0.168</b>	<b>0.906</b>	<b>0.778</b>

fingerprints as the sole drug representation, following the fingerprint-based pipeline in TEFDTA; only graph, which employs GNN-based molecular graphs alone, inspired by the design in GraphDTA; MACCS + graph, our proposed multiview drug fusion strategy that integrates molecular fingerprints and graphs; +residue-aware, which introduces DRN to incorporate residue-level modeling within the protein encoder; and +attention, the complete version of our model that further integrates multi-scale CNNs with attention-based residue reweighting for fine-grained protein representation.

As shown in Table 5, MACCS + graph consistently outperforms unimodal variants, confirming the effectiveness of the multiview drug representation. When the residue-aware module is incorporated, the performance improves markedly, indicating that residue-level modeling contributes to more discriminative protein features. Upon further integration of the attention mechanism, the model achieves the best overall performance, suggesting that attention enhances residue-level interactions by adaptively focusing on functionally relevant regions. This trend is also clearly illustrated in Fig. 4, showing that incorporating additional data modalities and structural representations accelerates convergence and reduces final loss. In addition, we evaluated Morgan (ECFP4) fingerprints in place of MACCS. Results are provided in Table S4, confirming competitive performance across standard fingerprint choices.

The introduction of the residue-aware module initially leads to slight fluctuations in the loss curve, likely due to the

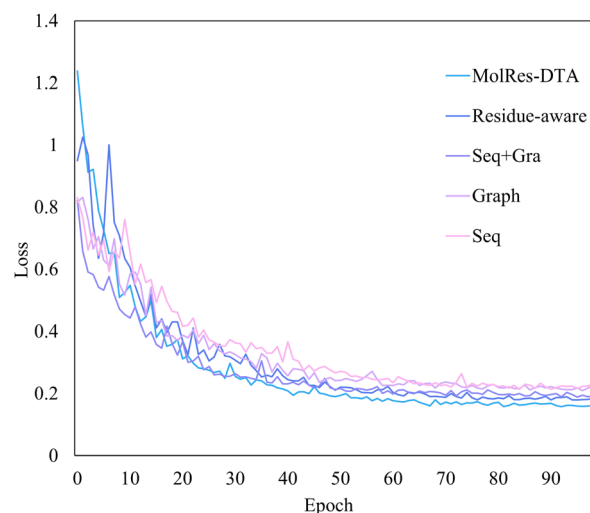


Fig. 4 Training loss curves of different model variants.



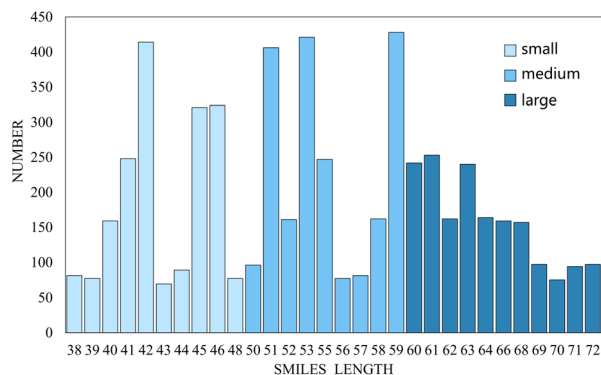


Fig. 5 SMILES length distribution in the Davis test set.

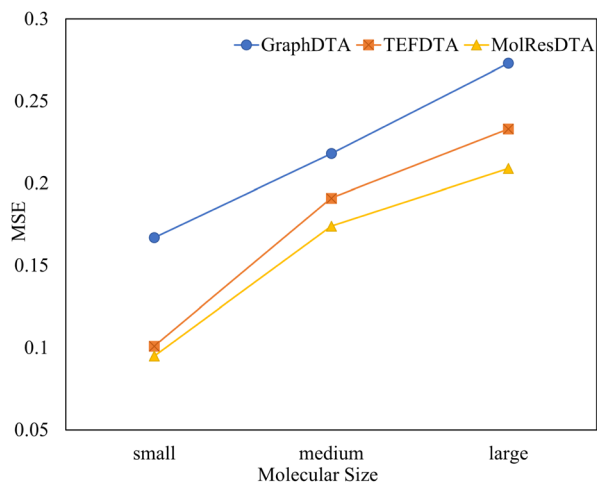


Fig. 6 Model performance across different molecular size groups.

additional learnable parameters and the adjustment process of residue-specific features. However, as training progresses, the model with residue-aware encoding consistently surpasses its baseline counterpart, confirming the long-term benefits of modeling residue-specific variation across protein sequences. Finally, the full MolRes-DTA model achieves both faster convergence and lower minimum loss, demonstrating the

synergistic advantages of multiview drug representation and residue-aware protein encoding while also confirming the overall stability and robustness of the model. To further validate the rationality of our model's design, we additionally performed an ablation study on various feature fusion strategies employed in MolRes-DTA, with results detailed in Table S5.

### 3.3 Role of multiview features in different molecular sizes

To investigate the adaptability of the proposed multiview drug feature fusion strategy to molecules of varying structural complexity, we adopt the length of SMILES strings as a practical proxy for structural complexity, enabling scalable and interpretable performance evaluation. Based on the distribution of SMILES lengths in the Davis test set (Fig. 5), we divided the drugs into three categories, small, medium, and large, using a tertile-based grouping strategy. Specifically, the lower and upper tertiles were used as cutoffs to assign 1859, 2079, and 1740 drugs to the small, medium, and large groups, respectively, with color gradients from light blue to dark blue denoting the size categories. Statistical analysis shows that the quantity of the three sets of data is roughly balanced. This observation suggests that the original dataset design likely aimed for a relatively uniform distribution of molecular sizes, which confirms the rationality of our classification method.

Based on this classification, we compared the model's performance on different molecular sizes with GraphDTA and TEFDTA. As illustrated in Fig. 6, all three models exhibit a general trend of increasing MSE as molecular size increases.

The performance of GraphDTA shows a linear degradation with the increase of molecular size. This suggests that its representational capacity becomes increasingly strained as the molecular graph grows in complexity. Additionally, its performance on small-sized compounds is also relatively poor. This aligns with our hypothesis that when molecular graphs are small, GNN-based models capture limited local structural information, and the absence of global semantic context further constrains predictive capability.

TEFDTA performs more robustly on small compounds, benefiting from its fingerprint-based representation that captures compact pharmacophoric patterns. However, its

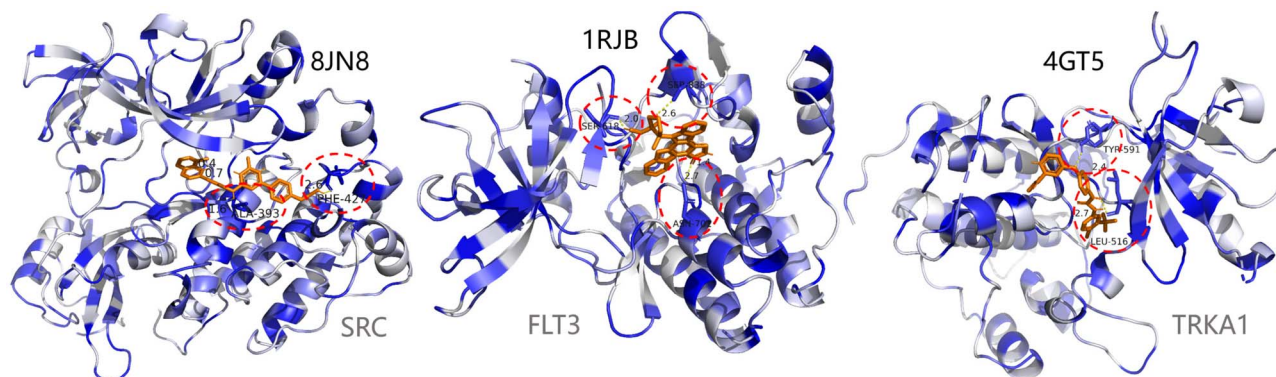


Fig. 7 Visual validation of DRN residue weights and molecular docking binding sites.



Table 6 External drug target pair affinity test

ID	Ligand	Target	Measured affinity	Predicted affinity
1033872	SARS-CoV-2 PLpro inhibitor	Replicase polyprotein 1ab	5.886	5.630
1370221	Acsmedchemlett	B-cell lymphoma 6 protein	6.876	5.762
46562	ALPRENOLOL	Beta-3 adrenergic receptor	6.932	6.784
430023	Benzenesulfonamide	Carbonic anhydrase 7	8.301	7.265
391242	ARC-3430	cAMP-dependent protein kinase catalytic subunit alpha	8.387	6.371

performance drops significantly for medium and large compounds. This decline reflects the limitations of its fixed-length fingerprint representation, which lacks the structural flexibility to scale with molecular complexity, resulting in the loss of critical topological information.

By contrast, MolRes-DTA consistently achieves lower MSE across all molecular size groups and exhibits particularly notable improvements in the medium and large subsets. These results demonstrate the effectiveness of our multiview fusion strategy, which integrates atom-level topological connectivity with global pharmacophoric context. This hybrid representation not only preserves TEFDTA's strength on small compounds but also introduces the structural adaptability required to model larger and more complex molecules, as shown in Table S6. The observed improvements affirm the robustness and generalizability of MolRes-DTA in addressing molecular size-dependent prediction challenges in DTA tasks.

To further validate these findings, we additionally analyzed molecular size effects on the KIBA dataset (see in Fig. S2 and Table S7), which encompasses a broader spectrum of protein families than Davis, and obtained consistent trends. Moreover, we explored alternative molecular size definitions, including molecular surface area, van der Waals volume, and molecular weight. Results from these analyses also support the conclusion that predictive accuracy decreases as molecular size increases, with the most pronounced differences observed between small and medium molecules. Detailed results are provided in Tables S8–S10.

However, the underlying mechanism driving this performance trend warrants further investigation. Although large molecules are relatively less frequent, the observed performance degradation cannot be fully explained by data imbalance alone. We further discuss that the likely reason is the inherently greater structural complexity and higher functional diversity of large molecules, which increases prediction difficulty. Overall, this consistent pattern across models underscores the significant impact of molecular size on prediction accuracy.

### 3.4 Model interpretability analysis

We conducted a structural interpretability analysis across representative drug-target pairs through weight visualization, molecular docking, and molecular dynamics simulations. Three protein–ligand complexes were randomly selected from the test set to cover low, medium, and high affinity ranges.

For analysis of weight visualization, the residue importance scores produced by DRN were mapped onto the corresponding experimental 3D structures for each target, and the resulting

spatial distributions were visualized using PyMOL. Residues were colored from white to blue to indicate increasing absolute weight.

Under this analysis pipeline, all three representative pairs (SRC-Dasatinib, FLT3-Lestaurtinib, and TRKA-Sorafenib) consistently exhibited higher DRN-assigned importance within residues that directly interact with the ligand, as determined by AutoDock Vina docking simulations (Fig. 7). This demonstrates that DRN effectively captures meaningful structural cues and highlights functionally relevant regions along the protein sequence. The predicted affinities alongside docking-derived binding estimates are summarized in Table S11, and the corresponding molecular dynamics simulations for these protein–ligand pairs are shown in Fig. S3.

### 3.5 Case study

We also incorporate affinity data from the latest release of BindingDB (an external source), to evaluate the generalization capability in real-world scenarios. We selected compound–target pairs that did not appear in either the training or test sets, covering diverse protein families including enzymes, kinases, GPCRs, and transcription factors. For each pair, we compared the reported binding affinity with MolRes-DTA's predicted value (Table 6).

Across these diverse protein families, MolRes-DTA generated affinity estimates that were generally consistent with experimental measurements. Most predictions fell within an acceptable deviation range, indicating that the model possesses robust generalization ability even when confronted with unseen targets.

## 4 Conclusions

In this study, we introduce MolRes-DTA, a novel DTA prediction model that integrates multiview drug encoding with a residue-aware representation to effectively address the molecular size effect and the varying contributions of residues. A comprehensive series of comparative experiments validate the superiority of our model, demonstrating the significance of its key components. Notably, the performance results showed that our model mitigates the accuracy decline observed in larger compounds. This finding provides valuable insights for drug design and development, particularly in enhancing the prediction of drug–target affinities for larger molecular entities.

Our work highlights the importance of integrating molecular graphs and fingerprints to improve molecular representation. Given that drug–target interactions occur within a dynamic



three-dimensional context, future research could benefit from incorporating molecular conformations and binding pocket information. Such an integration would enable richer structural and spatial representations, potentially leading to further improvements in prediction accuracy.

## Author contributions

Hongli Hou: data curation, methodology, investigation, writing – original draft; Qi Wei: methodology, writing – original draft; Dian Huang: supervision, writing – review and editing; Minglu Zhao: manuscript revision and data interpretation; Hongliang Duan: supervision and editing; Shengzhong Feng: supervision, writing – review and editing; all authors: validation and proofreading.

## Conflicts of interest

The authors declare no conflicts of interest.

## Data availability

The source code for model training, evaluation, and analysis is available at: <https://github.com/xiaohou-88/MolResDTA>. To ensure long-term accessibility and citation, an archived and versioned release of the repository has been deposited on Zenodo at: <https://doi.org/10.5281/zenodo.19029767>. The repository also includes links to the original datasets (Davis, KIBA, and BindingDB), as well as instructions for data preprocessing and full reproducibility of the experimental results.

Supplementary information (SI): additional experimental results. See DOI: <https://doi.org/10.1039/d5dd00365b>.

## Acknowledgements

This study is supported by the Young Scientists Fund-Type C of the National Natural Science Foundation of China (Grant No. 12404263) and Guangdong High Level Innovation Research Institute (2021B0909050004), and is approved by the Macao Polytechnic University (Submission ID: fca.e9b0.063c.b). The authors also appreciate the National Open Innovation Platforms for the New Generation Brain-inspired Artificial Intelligence established by Guangdong Institute of Intelligence Science and Technology.

## References

- 1 Y. Chu, A. C. Kaushik, X. Wang, W. Wang, Y. Zhang, X. Shan, D. R. Salahub, Y. Xiong and D.-Q. Wei, *Briefings Bioinf.*, 2021, **22**, 451–462.
- 2 D. Zhou, Z. Xu, W. Li, X. Xie and S. Peng, *Bioinformatics*, 2021, **37**, 4485–4492.
- 3 Y. Kalakoti, S. Yadav and D. Sundar, *ACS Omega*, 2022, **7**, 12138–12146.
- 4 L. Zhou, Z. Li, J. Yang, G. Tian, F. Liu, H. Wen, L. Peng, M. Chen, J. Xiang and L. Peng, *Molecules*, 2019, **24**, 1714.
- 5 W. Zhao, Y. Yu, G. Liu, Y. Liang, D. Xu, X. Feng and R. Guan, *Briefings Bioinf.*, 2024, **25**, bbae238.
- 6 N. Yousefi, M. Yazdani-Jahromi, A. Tayebi, E. Kolanthai, C. J. Neal, T. Banerjee, A. Gosai, G. Balasubramanian, S. Seal and O. Ozmen Garibay, *Briefings Bioinf.*, 2023, **24**, bbad136.
- 7 I. A. Guedes, C. S. de Magalhães and L. E. Dardenne, *Biophys. Rev.*, 2014, **6**, 75–87.
- 8 A. A. Naqvi, T. Mohammad, G. M. Hasan and M. I. Hassan, *Curr. Top. Med. Chem.*, 2018, **18**, 1755–1768.
- 9 V. Kairys, L. Baranauskiene, M. Kazlauskiene, D. Matulis and E. Kazlauskas, *Expert Opin. Drug Discovery*, 2019, **14**, 755–768.
- 10 A. Dhakal, C. McKay, J. J. Tanner and J. Cheng, *Briefings Bioinf.*, 2022, **23**, bbab476.
- 11 W. Shi, H. Yang, L. Xie, X.-X. Yin and Y. Zhang, *Health Inf. Sci. Syst.*, 2024, **12**, 30.
- 12 Q. Zhao, G. Duan, M. Yang, Z. Cheng, Y. Li and J. Wang, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 2022, **20**, 852–863.
- 13 J. Wang, X. Liu, S. Shen, L. Deng and H. Liu, *Briefings Bioinf.*, 2022, **23**, bbab390.
- 14 K.-K. Mak and M. R. Pichika, *Drug discovery today*, 2019, **24**, 773–780.
- 15 J. Zheng, X. Xiao and W.-R. Qiu, *Front. Genet.*, 2022, **13**, 859188.
- 16 S. Nag, A. T. Baidya, A. Mandal, A. T. Mathew, B. Das, B. Devi and R. Kumar, *3 Biotech*, 2022, **12**, 110.
- 17 H. Askr, E. Elgeldawi, H. Aboul Ella, Y. A. Elshaiar, M. M. Gomaa and A. E. Hassanien, *Artif. Intell. Rev.*, 2023, **56**, 5975–6037.
- 18 H. Öztürk, A. Özgür and E. Ozkirimli, *Bioinformatics*, 2018, **34**, i821–i829.
- 19 H. Öztürk, E. Ozkirimli and A. Özgür, *arXiv*, 2019, preprint, arXiv:1902.04166, DOI: [10.48550/arXiv.1902.04166](https://doi.org/10.48550/arXiv.1902.04166).
- 20 M. Gim, J. Choe, S. Baek, J. Park, C. Lee, M. Ju, S. Lee and J. Kang, *Bioinformatics*, 2023, **39**, i448–i457.
- 21 X. Zhu, J. Liu, J. Zhang, Z. Yang, F. Yang and X. Zhang, *Big Data Min. Anal.*, 2022, **6**, 1–10.
- 22 Z. Li, P. Ren, H. Yang, J. Zheng and F. Bai, *Bioinformatics*, 2024, **40**, btad778.
- 23 T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le and S. Venkatesh, *Bioinformatics*, 2021, **37**, 1140–1147.
- 24 L.-I. Tsui, T.-C. Hsu and C. Lin, *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2023, pp. 1–4.
- 25 Z. Zhu, Z. Yao, X. Zheng, G. Qi, Y. Li, N. Mazur, X. Gao, Y. Gong and B. Cong, *Comput. Biol. Med.*, 2023, **167**, 107621.
- 26 Z. Zhu, X. Zheng, G. Qi, Y. Gong, Y. Li, N. Mazur, B. Cong and X. Gao, *Expert Syst. Appl.*, 2024, **255**, 124647.
- 27 L. Zhao, H. Wang and S. Shi, *Bioinformatics*, 2024, **40**, btae594.
- 28 S. Chillistone and J. G. Hardman, *Anaesth. Intensive Care Med.*, 2017, **18**, 335–339.
- 29 B. Shin, S. Park, K. Kang and J. C. Ho, *Machine Learning for Healthcare Conference*, 2019, pp. 230–248.
- 30 Y. Zeng, X. Chen, Y. Luo, X. Li and D. Peng, *Briefings Bioinf.*, 2021, **22**, bbab117.



- 31 L. Zhang, W. Zeng, J. Chen, J. Chen and K. Li, *Biomed. Signal Process. Control*, 2024, **92**, 106110.
- 32 L. Chen, X. Tan, D. Wang, F. Zhong, X. Liu, T. Yang, X. Luo, K. Chen, H. Jiang and M. Zheng, *Bioinformatics*, 2020, **36**, 4406–4414.
- 33 Z. Zhu, Z. Yao, G. Qi, N. Mazur, P. Yang and B. Cong, *CAAI Trans. Intell. Technol.*, 2023, **8**, 1558–1577.
- 34 Z. Zhu, Y. Ding, G. Qi, B. Cong, Y. Li, L. Bai and X. Gao, *Eng. Appl. Artif. Intell.*, 2025, **147**, 110239.
- 35 H. Wu, J. Liu, T. Jiang, Q. Zou, S. Qi, Z. Cui, P. Tiwari and Y. Ding, *Neural Networks*, 2024, **169**, 623–636.
- 36 W. Tang, Q. Zhao and J. Wang, *IEEE Trans. Comput. Biol. Bioinf.*, 2025, 1–12.
- 37 M. I. Davis, J. P. Hunt, S. Herrgard, P. Ciceri, L. M. Wodicka, G. Pallares, M. Hocker, D. K. Treiber and P. P. Zarrinkar, *Nat. Biotechnol.*, 2011, **29**, 1046–1051.
- 38 J. Tang, A. Szwajda, S. Shakyawar, T. Xu, P. Hintsanen, K. Wennerberg and T. Aittokallio, *J. Chem. Inf. Model.*, 2014, **54**, 735–743.
- 39 T. Liu, Y. Lin, X. Wen, R. N. Jorissen and M. K. Gilson, *Nucleic Acids Res.*, 2007, **35**, D198–D201.
- 40 A. C. Nascimento, R. B. Prudêncio and I. G. Costa, *BMC Bioinf.*, 2016, **17**, 1–16.
- 41 T. He, M. Heidemeyer, F. Ban, A. Cherkasov and M. Ester, *J. Cheminf.*, 2017, **9**, 1–14.
- 42 K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J. B. Ghasemi and A. Masoudi-Nejad, *Bioinformatics*, 2020, **36**, 4633–4642.

