JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

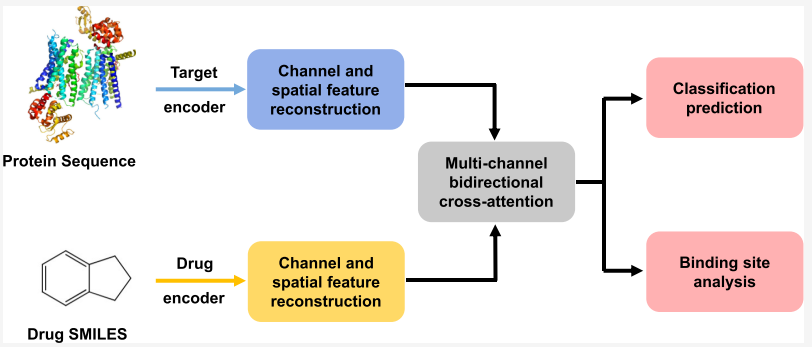
pubs.acs.org/jcim Article

ReduMixDTI: Prediction of Drug−Target Interaction with Feature Redundancy Reduction and Interpretable Attention Mechanism

Mingqing Liu, Xuechun Meng, Yiyang Mao, Hongqi Li,* and Ji Liu*

Cite This: J. Chem. Inf. Model. 2024, 64, 8952−8962 Read Online

ACCESS | Metrics & More | Article Recommendations



ABSTRACT: Identifying drug−target interactions (DTIs) is essential for drug discovery and development. Existing deep learning approaches to DTI prediction often employ powerful feature encoders to represent drugs and targets holistically, which usually cause significant redundancy and noise by neglecting the restricted binding regions. Furthermore, many previous DTI networks ignore or simplify the complex intermolecular interaction process involving diverse binding types, which significantly limits both predictive ability and interpretability. We propose ReduMixDTI, an end-to-end model that addresses feature redundancy and explicitly captures complex local interactions for DTI prediction. In this study, drug and target features are encoded by using graph neural networks and convolutional neural networks, respectively. These features are refined from channel and spatial perspectives to enhance the representations. The proposed attention mechanism explicitly models pairwise interactions between drug and target substructures, improving the model's understanding of binding processes. In extensive comparisons with seven state-of-the-art methods, ReduMixDTI demonstrates superior performance across three benchmark data sets and external test sets reflecting real-world scenarios. Additionally, we perform comprehensive ablation studies and visualize protein attention weights to enhance the interpretability. The results confirm that ReduMixDTI serves as a robust and interpretable model for reducing feature redundancy, contributing to advances in DTI prediction.

INTRODUCTION

Drugs alter the structure and function of biological complexes by recognizing and binding to specific regions on their targets, a process known as drug-target interaction (DTI).[1] Typically, drugs refer to small molecular compounds or tiny peptides, while the majority of targets are proteins.[2] The identification of DTI is crucial for predicting drug side effects and exploring drug repositioning, which aims to identify new drugs or novel targets for existing drugs.[3−5] However, the experimentally validated DTI data are far fewer than the potential drug-target combinations, limiting progress in drug research and development.[6] Although wet-lab methods for DTI identification are considered reliable, they cannot be applied on a large scale due to being time-consuming, expensive, and labor-intensive.[7,8]

In contrast, computational methods have gained attention for their efficiency in identifying potential DTIs. These methods can guide in vitro validation and significantly reduce experimental time and costs.[9] Traditional computational approaches primarily use molecular docking[10] and virtual screening.[11] Molecular docking depends on available 3-dimensional (3D) protein structures, which are much fewer than the 1-dimensional (1D) protein sequences, limiting its large-scale applications.[12] Virtual screening predicts new molecules based on known active molecules for a given protein, but its performance suffers when sufficient active molecules are unavailable.[13]

Recently, deep learning methods with powerful feature extraction capabilities have achieved significant advancements in handling discrete symbolic data. For DTI prediction, proteins

Received: August 28, 2024
Revised: October 30, 2024
Accepted: November 5, 2024
Published: November 21, 2024

https://doi.org/10.1021/acs.jcim.4c01554
J. Chem. Inf. Model. 2024, 64, 8952−8962

can be represented as 1D amino acid sequences, while drugs can be encoded using Simplified Molecular-Input Line-Entry System (SMILES) sequences.[14] Consequently, several end-to-end approaches have been developed for DTI prediction.[15−17] These data-driven deep learning methods enable large-scale validation within a relatively short time and reduce the reliance on expert-defined features.

Existing DTI prediction methods generally consist of two main components: the extraction of features from drugs and proteins and the modeling of their interactions. Previous studies have extensively explored drug and protein feature representations. For instance, Ozturk et al.[18] proposed the DeepDTA model based on convolutional neural networks (CNNs) to extract biological information from 1D protein sequences and drug SMILES. Subsequently, Tsubaki et al.[19] utilized 2-dimensional (2D) molecular graphs to represent drugs and extracted features using a graph convolution network (GCN) model, demonstrating the potential of graph models for molecular feature extraction. Inspired by DeepDTA, Abbasi et al.[20] proposed DeepCDA, incorporating long short-term memory (LSTM) units after each CNN layer to further model sequence associations. Furthermore, Huang et al.[21] and Zhang et al.[22] employed Transformer encoders[23] to learn drug and protein features, with Zhang et al.[24] later enhancing feature representation by incorporating pharmacophore information. Zhu et al.[25] constructed a variational autoencoder combining a cascade of Transformers and CNNs to achieve single-molecule-level feature representation of drugs and targets.

For pairwise representation, early studies often simplified the interaction process by concatenating drug and protein features for DTI prediction.[18] More recent research has focused on modeling the binding process between drugs and proteins. For instance, Chen et al.[26] proposed TransformerCPI, using a one-side Transformer decoder to model interactions and highlight important protein substructures for molecules. Huang et al.[21] designed a pairwise interaction module based on CNN to capture semantic relationships among substructures. Zhao et al.[16] introduced HyperAttentionDTI, employing fully connected (FC) layers to transform drug and protein representations into attention matrices. Additionally, DrugBAN[17] treated DTI prediction as a multimodal fusion problem, capturing local structural relationships with a bilinear attention mechanism.[27] Additionally, Zhu et al.[28,29] further leverage multiscale information from proteins and drugs, enabling the capture of interaction features at different granularities.

Although the above methods have made significant improvements in the field of DTI or DTA prediction, there are still some challenges. Existing DTI prediction methods often rely on deep learning models to automatically extract features from drugs and targets based on their complete information.[18−21] However, they often ignore the fact that the binding processes between drugs and proteins occur in local regions rather than throughout entire structures. Using full biomolecular features inevitably results in redundant information for DTI prediction, even introducing noise and increasing computational costs. Additionally, the drug recognition of targets is driven by chemical interactions between specific biomolecular substructures. These interactions involve a complex array of binding forces, such as ionic bonds, hydrogen bonds, and hydrophobic effects, all of which contribute to the formation of stable drug-target complexes.[30−32] Nevertheless, most DTI prediction models ignore or simplify these intricate interaction processes, thereby limiting their predictive power.

To address these fundamental questions, we proposed an improved end-to-end deep learning model named ReduMixDTI, designed to reduce redundancy in global feature representations while effectively modeling local interactions with various binding forces. To this end, protein sequences and drug molecular graphs are first encoded by CNN and GCN, respectively. We then introduce a plug-and-play channel and spatial feature reconstruction (CSFR) module that dynamically adjusts channel importance to highlight relevant information for DTI prediction while reorganizing spatial content for more informative representations. Following this, a multichannel bidirectional cross-attention (MBCA) module is employed to explicitly model various types of interactions between substructures and capture drug-target pairwise representations. To evaluate the model's performance, we compare ReduMixDTI with seven other state-of-the-art (SOTA) DTI prediction models. Finally, we visualize the predicted important regions in 3D drug−protein complexes and compare them to actual binding sites.

In summary, the main contributions of our work include

- We propose an end-to-end framework called ReduMixDTI to reduce redundant information in feature extraction and represent complex interaction processes for DTI prediction. ReduMixDTI incorporates a feature redundancy reduction module that integrates important features of drugs and targets from multiple perspectives, along with an interaction modeling module based on a multichannel bidirectional cross-attention mechanism.
- We introduce a plug-and-play module named channel and spatial feature reconstruction. This module adaptively weights features extracted by drug and protein encoders using multiscaled channel attention. Additionally, the spatial information in feature maps is reorganized by segmenting based on information content, enhancing the representation of drug and target features.
- We propose a multichannel bidirectional cross-attention module to explicitly model different interaction dependencies between drug and target substructures. Moreover, this module captures various types of binding forces involved with the same substructures through different attention heads. A subsequent linear layer combines the multihead outputs to adaptively integrate these complex interaction patterns, yielding more realistic and powerful features of drug-target pairs.

MATERIALS AND METHODS

Data Sets. To comprehensively evaluate our proposed model and other SOTA baselines, we select three public DTI data sets: Human, BindingDB, and BioSNAP, as described in Table 1. Specifically, the balanced Human data set was constructed by Liu et al.,[33] who collected positive samples from the DrugBank database 4.1[34] and used an in silico screening method to obtain highly credible negative samples of equal size. Gao et al.[35] constructed the BindingDB data set, which was later refined into a low-bias version by Bai et al.[36]

Table 1. Summary of the Three Benchmark Data Sets

data set	interaction	drug	protein
BindingDB	49199	14643	2623
Human	6728	1052	852
BioSNAP	27464	4510	2181

https://doi.org/10.1021/acs.jcim.4c01554
J. Chem. Inf. Model. 2024, 64, 8952−8962

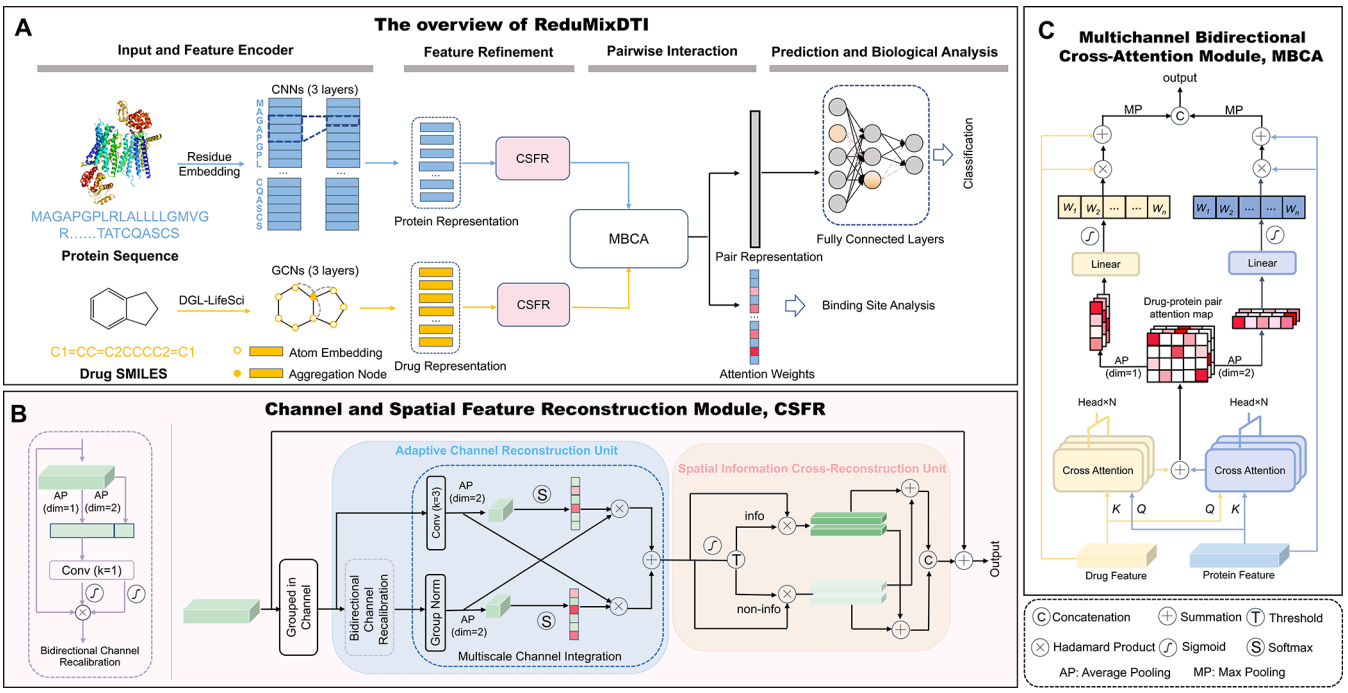


Figure 1. Architecture of ReduMixDTI. (A) Box shows the overall pipeline from drug SMILES and protein sequence inputs to DTI prediction results. (B) Channel and spatial feature reconstruction (CSFR) module, part of ReduMixDTI, is a feature redundancy removal module containing two units: the adaptive channel reconstruction unit and the spatial information cross-reconstruction unit. Drug and protein representations pass through these units to reduce the feature redundancy and enhance important feature representations. (C) Multichannel bidirectional cross-attention (MBCA) module learns pairwise representations based on local interactions between refined drug and protein features for DTI prediction.

They removed drugs or proteins that appeared exclusively in either positive or negative samples. For the BioSNAP data set, Zitnik et al.[37] collected positive samples, while negative samples were randomly generated by pairing drugs and proteins not appearing in the positive samples.

Framework of ReduMixDTI. The overall pipeline of the ReduMixDTI architecture is shown in Figure 1A, consisting of three key parts: Input and Feature Encoder, Individual Feature Refinement, and Pairwise Feature Interaction. Following these, the Prediction and Biological Analysis component includes FC layers for DTI prediction and attention weight visualization for interpretable analysis. The remainder of this section describes each component of ReduMixDTI in detail.

Input and Feature Encoder. The Input and Feature Encoder component processes drug SMILES inputs and 1D protein sequence inputs, respectively. For drug compounds, each SMILES string is converted into a 2D molecular graph using the DGL-LifeSci package,[38] where nodes represent atoms, and edges correspond to atomic bonds. Node information is initialized with nine attributes: eight chemical properties predefined by the DGL-LifeSci package and one binary label indicating whether the atom is virtual. Consequently, the drug embedding matrix is denoted as $\mathbf{E}_d \in \mathbb{R}^{\theta_d \times 75}$, where θ_d is the maximum allowed number of nodes, and 75 is the dimension of each node's integer vector. To obtain a dense matrix of the drug $\mathbf{X}_d \in \mathbb{R}^{\theta_d \times C_d}$, a simple linear transformation maps each node vector to C_d dimensions.

Next, a three-layer GCN block effectively learns the graph representation, incorporating skip connections after each layer to prevent information loss. Each atom feature vector is updated by aggregating features from itself and neighborhood atoms, capturing molecular substructure representations at various scales. This node-level drug representation method facilitates learning local interactions with protein fragments in the subsequent Pairwise Feature Interaction component. The drug encoder is described by the following equation:

$$\mathbf{H}_d^{(l+1)} = \sigma(\mathbf{D}^{-1/2}\hat{\mathbf{A}}\mathbf{D}^{-1/2}\mathbf{H}_d^{(l)}\mathbf{W}_g^{(l)}) \quad (1)$$

where $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ is the adjacency matrix including the self-connection, $\mathbf{I}$ is the identity matrix, $\mathbf{D}$ is the diagonal node degree matrix of $\hat{\mathbf{A}}$, $\mathbf{H}_d^{(l)} \in \mathbb{R}^{\theta_d \times C_d}$ is the latent representation of the lth layer with $\mathbf{H}_d^{(0)} = \mathbf{X}_d$, $\mathbf{W}_g^{(l)}$ is the layer-specific learnable matrix, and $\sigma(\cdot)$ is the rectified linear unit (ReLU) activation function.

For protein sequences, each amino acid is first mapped as an embedding vector using a learnable embedding matrix $Emb \in \mathbb{R}^{24 \times C_p}$, where C_p denotes the embedding dimensionality and 24 represents the number of amino acid types. By looking up Emb, we obtain the initialized protein feature matrix $\mathbf{X}_p \in \mathbb{R}^{\theta_p \times C_p}$, where θ_p is the maximum allowed length of the protein sequence.

The protein feature encoder consists of a three-layer 1D CNN block with batch normalization, which extracts local residue patterns in proteins as the convolution filter slides over $\mathbf{X}_p$. The protein encoder is written as

$$\mathbf{H}_p^{(l+1)} = \sigma(\mathrm{BN}(\mathrm{CNN}(\mathbf{W}_c^{(l)}, \mathbf{H}_p^{(l)})) \quad (2)$$

where $W_c^{(l)}$ is the learnable weight matrix in the lth CNN layer, $\mathbf{H}_p^{(l)}$ denotes the hidden protein representation at the lth layer with $\mathbf{H}_p^{(0)} = X_p$, and BN indicates batch normalization. After the GCN and CNN encoders, we obtain the latent global representations $\mathbf{H}_d^{(3)} \in \mathbb{R}^{L_d \times C_d}$ for drugs and $\mathbf{H}_p^{(3)} \in \mathbb{R}^{L_p \times C_p}$ for proteins, where L_d and L_p denote the number of encoded substructures in a drug and a protein, respectively.

https://doi.org/10.1021/acs.jcim.4c01554
J. Chem. Inf. Model. 2024, 64, 8952−8962

Individual Feature Refinement. We design the CSFR module by inheriting the EMA module[39] and the SCConv module.[40] This module aims to extract the most relevant information from both channel and spatial perspectives while suppressing global features that are irrelevant to the DTI prediction. As illustrated in Figure 1B, the CSFR module consists of two units: the adaptive channel reconstruction (ACR) unit and the spatial information cross-reconstruction (SICR) unit.

The ACR unit utilizes channel redundancy across different directions and scales to enhance the feature representation. Concretely, the encoding output is grouped as $\mathbf{H}_g \in \mathbb{R}^{G\times L\times C//G}$, where G denotes the number of groups. $\mathbf{H}_g$ undergoes average pooling in both horizontal and vertical directions to extract channel descriptors, indicating channel interdependencies across different dimensions of the feature space. These descriptors are concatenated and passed through a 1D CNN with a kernel size of 1 to integrate information from both directions. The Sigmoid function is utilized to produce attention weights, subsequently applied to reweight $\mathbf{H}_g$, enabling adaptive learning of channel importance based on complementary direction information.

Next, the ACR unit explores channel redundancy across different scales. At the first scale, the grouped feature $\mathbf{H}_g$ is fed into a 1D CNN with a kernel size of 3 to produce $\mathbf{F}_1$, where the enlarged convolutional receptive field captures broader contextual representations. At the second scale, the channel-tuned output undergoes grouped normalization to produce $\mathbf{F}_2$, representing the foundational information. Both $\mathbf{F}_1$ and $\mathbf{F}_2$ are passed through global average pooling and converted into channel attention weights $\mathbf{A}_1$ and $\mathbf{A}_2$ via the Softmax function. Cross-multiplication and summation of the multiscaled representation matrices with their respective channel importance weights dynamically adjust features and integrate information. This process further aggregates channel importance from different scales, highlighting the most critical features based on a comprehensive understanding of diverse contextual information:

$$\mathbf{H}^c = \mathbf{F}_1 \otimes \mathbf{A}_2 + \mathbf{F}_2 \otimes \mathbf{A}_1 \tag{3}$$

where $\mathbf{H}^c$ is the output of the ACR unit, and $\otimes$ denotes the Hadamard product.

Finally, the SICR unit addresses spatial redundancy by distinguishing informative feature maps from less informative features based on spatial content. Each element in $\mathbf{H}^c$ is scaled by the Sigmoid function to a range of 0−1 and then gated by a threshold. Values above the threshold form the informative mask M_1, while those below form the noninformative mask M_2 (the threshold is set to 0.3 in the experiments). Element-wise multiplication is applied between $\mathbf{H}^c$ and both M_1 and M_2 separately to achieve information segmentation, resulting in the informative feature matrix $\mathbf{H}_1^M = M_1 \otimes \mathbf{H}^c$ and the less informative one $\mathbf{H}_2^M = M_2 \otimes \mathbf{H}^c$. The matrices $\mathbf{H}_1^M$ and $\mathbf{H}_2^M$ are then concatenated using a cross-summation method, denoted as

$$\begin{cases} \mathbf{H}_1^M = \mathbf{H}_{11}^M \oplus \mathbf{H}_{22}^M, \\ \mathbf{H}_2^M = \mathbf{H}_{21}^M \oplus \mathbf{H}_{12}^M, \\ \mathbf{H}^M = \mathbf{H}_1^M \cup \mathbf{H}_2^M. \end{cases} \tag{4}$$

where $\oplus$ denotes element-wise summation, $\cup$ represents concatenation, and $\mathbf{H}^M$ is the enhanced drug or protein representation, with $\mathbf{H}_d^M \in \mathbb{R}^{L_d\times C_d}$ and $\mathbf{H}_p^M \in \mathbb{R}^{L_p\times C_p}$.

Pairwise Feature Interaction. To effectively model localized molecular interactions, we developed the MBCA module. Chemical bonds between substructures can form through various mechanisms. For instance, a carboxyl group on a drug has the potential to form ionic bonds with positively charged amino acid residues in proteins or act as a hydrogen bond acceptor. Similarly, the active site of a protein, containing multiple polar and charged amino acid residues, may interact with a positively charged amino group in the drug molecule through ionic bonding or form hydrogen bonds with hydroxyl groups. As shown in Figures 1C and 2, the multihead mechanism

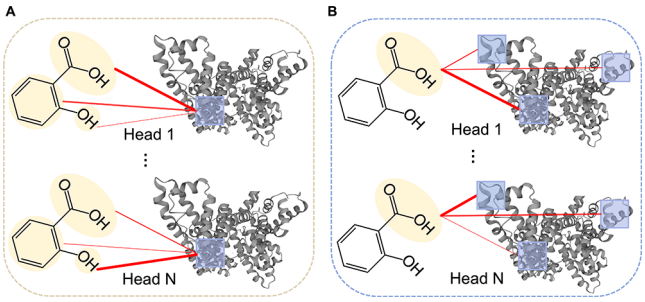


Figure 2. Multihead mechanism captures various potential binding patterns between a biomolecular substructure and its corresponding counterpart. (A) Protein active site interacts with various drug functional groups, resulting in distinct binding interactions across different heads. (B) Drug functional group interacts with different protein substructures, forming distinct binding interactions across different heads.

in MBCA projects drug and protein features into multiple subspaces, highlighting potential interaction patterns across different substructures and capturing various binding possibilities.

The bidirectional cross-attention mechanism in MBCA, similar to cross-attention in Transformers,[23] requires input queries $\mathbf{Q}$ and keys $\mathbf{K}$. Drug features $\mathbf{H}_d^M$ and protein features $\mathbf{H}_p^M$ are alternately assigned as $\mathbf{Q}$ and $\mathbf{K}$, allowing similarity computation between each drug substructure and all protein substructures to assess binding potential and vice versa. This approach effectively captures fine-grained interactions between drug and protein substructures, formulated as follows.

$$\mathbf{Q}_i = \mathbf{Q}W_i^Q, \quad \mathbf{K}_i = \mathbf{K}W_i^K, \quad i = 1, \cdots, h \tag{5}$$

$$\mathbf{A}_i = Softmax\left(\frac{\mathbf{Q}_i\mathbf{K}_i^T}{\sqrt{dim_k}}\right), \quad i = 1, \cdots, h \tag{6}$$

where $\mathbf{A}_i$, $\mathbf{Q}_i$, and $\mathbf{K}_i$ represent the ith head of the cross-attention result, query, and key matrices, respectively. Here, $\mathbf{A}_i \in \mathbb{R}^{L_Q\times L_K}$, the multihead attention $\mathbf{A} \in \mathbb{R}^{h\times L_Q\times L_K}$, h denotes the number of heads, L_Q the length of $\mathbf{Q}$, and L_K the length of $\mathbf{K}$. The *Softmax* function normalizes values to a probability distribution, with dim_k as a scaling factor.

After multihead bidirectional cross-attention, matrix $\mathbf{A}_{dp}$ represents the probability distribution map where each drug substructure evaluates its binding potential with protein substructures, while A_{pd} captures the reverse binding potential. The MBCA module combines these bidirectional attention maps into a comprehensive matrix $\mathbf{A} = \mathbf{A}_{dp} + \mathbf{A}_{pd}$, $\mathbf{A} \in \mathbb{R}^{h\times L_d\times L_p}$, aggregating crucial binding region information on drug-target complexes. Subsequently, average pooling is applied to $\mathbf{A}$ along both the L_d and L_p dimensions, creating global descriptors for

https://doi.org/10.1021/acs.jcim.4c01554
J. Chem. Inf. Model. 2024, 64, 8952−8962

drugs and proteins that reflect the binding-related substructure weight distribution derived from the complex.

Since each attention head captures different interaction aspects, a linear transformation is applied to global descriptors of drugs and proteins. This adaptive integration of patterns from each head facilitates the representation of complex interactions. A Sigmoid activation is then applied to refine the descriptors, yielding interaction weights $\mathbf{A}_d \in \mathbb{R}^{L_d \times C_d}$ and $\mathbf{A}_p \in \mathbb{R}^{L_p \times C_p}$, updating drug and protein features $\mathbf{F}_d$ and $\mathbf{F}_p$ as follows:

$$\mathbf{F}_d = \mathbf{H}_d^M + \mathbf{H}_d^M \otimes \mathbf{A}_d \tag{7}$$

$$\mathbf{F}_p = \mathbf{H}_p^M + \mathbf{H}_p^M \otimes \mathbf{A}_p \tag{8}$$

Finally, $\mathbf{F}_d$ and $\mathbf{F}_p$ undergo max pooling and concatenation, forming a joint representation of the drug-target pair.

To compute the DTI probability, we utilize FC layers with leaky ReLU activation and Dropout. Given that we treat DTI prediction as a binary classification task, the cross-entropy loss $\mathcal{L}$ is employed to train our model:

$$\mathcal{L} = -[y\log(\hat{y}) + (1 - y)\log(1 - \hat{y})] \tag{9}$$

where y is the binary interaction label, and $\hat{y}$ is the predicted likelihood.

EXPERIMENTS AND RESULTS

Evaluation Criteria. In our study, we adopt the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC) as the major metrics to assess the model's performance. Additionally, we calculate accuracy, recall and precision, specificity, and F1 score as supplementary metrics, which are formulated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{10}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{12}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{13}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

where TP and TN denote the numbers of positive and negative samples predicted correctly, while FP and FN are the numbers of positive and negative samples predicted incorrectly.

In the experiment, the data sets were randomly split into training, validation, and test sets in a ratio of 7:1:2. For evaluation, we report the mean and standard deviation (std) of each metric across five independent runs to ensure stable results.

Implementation Details. ReduMixDTI is implemented in Python 3.7 and Pytorch 1.9.0, utilizing an NVIDIA GeForce RTX 3090 GPU with 12GB of memory. The model is trained for up to 100 epochs by using the Adam optimizer, with an early stop method employed to mitigate overfitting. The model with the highest AUROC on the validation set is selected for final evaluation on the test set. To further optimize the performance of ReduMixDTI, we utilized a grid search to configure hyperparameters. The specific configurations are summarized in Table 2.

Table 2. Hyperparameter Configuration for ReduMixDTI

parameters	value
number of filters in protein CNN encoder	[128, 128, 128]
kernel size in protein CNN encoder	[3, 6, 9]
hidden node dimensions in drug GCN encoder	[128, 128, 128]
gate threshold of feature spatial information	0.3
number of attention heads	8
dropout rate	0.1
batch size	32
learning rate	5e−5
epoch	100

Baseline Models. To evaluate the performance of our proposed ReduMixDTI model, we compared it against seven SOTA baselines in the DTI prediction. These methods were selected for their diverse feature extraction techniques and interaction modeling strategies, allowing for a comprehensive evaluation of our model's effectiveness.

The specific methods include (1) TransformerCPI[26] employs a pretrained model for protein embeddings and uses Transformer encoder layers[23] to capture drug−protein interactions. (2) MolTrans[21] represents drug and protein substructure features with Transformer architecture and uses a CNN-based module to learn interactions. (3) GraphDTA[41] explores four different graph neural networks to learn drug molecular graphs and employs CNN to extract protein features. The drug and protein representations are then concatenated directly for subsequent prediction. Notably, GraphDTA was initially designed for a regression task, so the output of its last FC layer is modified to two classes, and cross-entropy loss is used for classification. (4) HyperAttentionDTI[16] uses CNN blocks for drug and protein feature encoding, with learned biological representations fed into FC layers to model interactions. (5) DrugBAN[17] treats DTI prediction as a multimodal learning task and uses a bilinear attention network to encode pairwise interaction. (6) MGNDTI[42] uses the large language model RetNet[43] for feature extraction from drug and protein sequences, integrating molecular graph information using a GCN encoder. A gating network combines multimodal information to represent drug-target pairs. (7) BINDTI[44] combines CNN with Transformer self-attention layers[23] for protein features, while GCN layers are used for drugs, and employs multihead self-attention with the Intention network[45] to model interactions. All methods are tested with hyperparameters as suggested in their respective original papers.

Performance Comparison with Baselines. The comparison results on three benchmark data sets are presented in Table 3. ReduMixDTI consistently achieves the highest performance in terms of AUROC, AUPRC, accuracy, precision, specificity, and F1, while achieving the best recall on the BindingDB and Human data sets and the second-best recall on the BioSNAP data set. For the BindingDB data set, we observe improvements in AUROC of 1.0% (from 0.960 to 0.970) and in AUPRC of 1.2% (from 0.949 to 0.961) over the best-performing baseline model. Similarly, for the BioSNAP data set, AUROC and AUPRC improve by 1.2% (from 0.911 to 0.923) and 1.1% (from 0.917 to 0.928), respectively. The improvements on the Human data set are 0.3% for AUROC (from 0.986 to 0.989) and 0.3% for AUPRC (from 0.987 to 0.990). These results indicate the improved performance of the proposed ReduMixDTI model in DTI prediction.

https://doi.org/10.1021/acs.jcim.4c01554
J. Chem. Inf. Model. 2024, 64, 8952−8962

Table 3. Models Performance Comparison on the BindingDB, BioSNAP, and Human Data Sets (Statistics Over Five Random Runs)[a]

model	AUROC (std)	AUPRC (std)	accuracy (std)	recall (std)	precision (std)	specificity (std)	F1 (std)
BindingDB data set							
TransformerCPI	0.929 (0.001)	0.924 (0.003)	0.878 (0.003)	0.870 (0.013)	0.856 (0.015)	0.881 (0.015)	0.862 (0.007)
MolTrans	0.944 (0.009)	0.932 (0.011)	0.890 (0.014)	0.891 (0.016)	0.864 (0.018)	0.865 (0.019)	0.871 (0.017)
GraphDTA	0.947 (0.002)	0.935 (0.003)	0.893 (0.002)	0.886 (0.012)	0.873 (0.008)	0.885 (0.013)	0.879 (0.004)
HyperAttentionDTI	*0.960 (0.001)*	*0.949 (0.002)*	0.907 (0.002)	*0.910 (0.005)*	0.868 (0.005)	0.903 (0.004)	0.887 (0.001)
DrugBAN	0.955 (0.001)	0.945 (0.002)	*0.910 (0.003)*	0.897 (0.005)	*0.895 (0.005)*	*0.911 (0.004)*	*0.895 (0.002)*
MGNDTI	0.952 (0.003)	0.935 (0.005)	0.893 (0.004)	0.899 (0.001)	0.855 (0.008)	0.885 (0.006)	0.876 (0.002)
BINDTI	0.956 (0.001)	0.943 (0.001)	0.901 (0.003)	0.904 (0.009)	0.868 (0.010)	0.899 (0.010)	0.886 (0.001)
ReduMixDTI	**0.970 (0.003)**	**0.961 (0.002)**	**0.918 (0.006)**	**0.915 (0.010)**	**0.901 (0.012)**	**0.920 (0.011)**	**0.907 (0.008)**
BioSNAP data set							
TransformerCPI	0.868 (0.004)	0.879 (0.006)	0.796 (0.001)	0.799 (0.007)	0.797 (0.005)	0.787 (0.004)	0.798 (0.001)
MolTrans	0.896 (0.005)	0.901 (0.004)	0.820 (0.012)	0.841 (0.026)	0.809 (0.027)	0.793 (0.024)	0.824 (0.009)
GraphDTA	0.889 (0.003)	0.891 (0.004)	0.827 (0.006)	0.838 (0.010)	0.827 (0.013)	0.823 (0.012)	0.831 (0.007)
HyperAttentionDTI	*0.911 (0.004)*	*0.917 (0.004)*	0.830 (0.005)	**0.882 (0.010)**	0.801 (0.009)	0.819 (0.008)	*0.840 (0.004)*
DrugBAN	0.907 (0.005)	0.910 (0.011)	*0.831 (0.004)*	0.840 (0.011)	*0.834 (0.010)*	0.820 (0.011)	0.836 (0.003)
MGNDTI	0.901 (0.003)	0.902 (0.004)	0.826 (0.007)	0.852 (0.007)	0.810 (0.013)	0.821 (0.010)	0.830 (0.006)
BINDTI	0.896 (0.003)	0.894 (0.004)	0.829 (0.006)	0.830 (0.011)	0.827 (0.011)	*0.829 (0.015)*	0.828 (0.005)
ReduMixDTI	**0.923 (0.003)**	**0.928 (0.002)**	**0.848 (0.003)**	*0.871 (0.011)*	**0.836 (0.010)**	**0.832 (0.011)**	**0.853 (0.006)**
Human data set							
TransformerCPI	0.973 (0.004)	0.973 (0.004)	0.923 (0.009)	0.934 (0.007)	0.920 (0.003)	0.921 (0.008)	0.927 (0.004)
MolTrans	0.981 (0.003)	0.983 (0.003)	0.943 (0.003)	0.932 (0.014)	0.950 (0.013)	*0.945 (0.013)*	0.941 (0.003)
GraphDTA	0.980 (0.002)	0.979 (0.004)	0.941 (0.008)	0.937 (0.011)	0.925 (0.013)	0.943 (0.015)	0.931 (0.006)
HyperAttentionDTI	*0.986 (0.001)*	0.986 (0.001)	0.937 (0.010)	0.949 (0.011)	0.923 (0.017)	0.934 (0.012)	0.936 (0.008)
DrugBAN	0.982 (0.004)	0.980 (0.006)	0.938 (0.005)	0.939 (0.015)	0.935 (0.011)	0.938 (0.018)	0.937 (0.006)
MGNDTI	0.985 (0.002)	*0.987 (0.002)*	*0.949 (0.007)*	*0.951 (0.017)*	*0.950 (0.010)*	0.941 (0.012)	*0.950 (0.007)*
BINDTI	0.982 (0.003)	0.983 (0.003)	0.944 (0.007)	0.948 (0.013)	0.943 (0.013)	0.939 (0.015)	0.945 (0.007)
ReduMixDTI	**0.989 (0.001)**	**0.990 (0.001)**	**0.953 (0.006)**	**0.952 (0.008)**	**0.953 (0.010)**	**0.953 (0.010)**	**0.952 (0.008)**

[a]Note: The bold value corresponds to the best performance method for each metric, while the italized value denotes the second-best performance method.

Table 4. Statistical Tests on BindingDB, BioSNAP, Human Data Sets Over Five Random Runs

	BindingDB		BioSNAP		human	
	t-value	*p*-value	*t*-value	*p*-value	*t*-value	*p*-value
TransformerCPI	28.3509	<0.001	18.0036	<0.001	6.3356	<0.005
MolTrans	4.9209	<0.01	12.5442	<0.001	10.3983	<0.001
GraphDTA	16.5596	<0.001	15.5543	<0.001	8.0749	<0.005
HyperAttentionDTI	5.6074	<0.005	5.3075	<0.01	6.9637	<0.005
DrugBAN	7.0486	<0.005	6.1598	<0.005	5.6514	<0.005
MGNDTI	8.2129	<0.005	6.9625	<0.005	4.6744	<0.01
BINDTI	10.8663	<0.001	8.9937	<0.001	5.2761	<0.01

In addition, statistical tests were conducted to evaluate the significance of the improvements achieved by ReduMixDTI compared to each baseline model. Table 4 presents the *t* test results for the AUROC values obtained from five random data splits. Higher t-values correspond to lower p-values, with p-values less than 0.05 typically indicating statistical significance. The *t* test results confirm that ReduMixDTI substantially outperforms baseline models on the BindingDB, BioSNAP, and Human data sets.

The promising performance of ReduMixDTI can be attributed to two main aspects. First, it reduces channel and spatial redundancy in the holistic feature representations of drugs and proteins, enhancing the contextual understanding derived from multiscaled convolutional operations. Second, the model employs a bidirectional cross-attention strategy to focus on local interactions and extend the distinct binding possibilities captured by the multihead mechanism. This attention strategy enhances the model's understanding of the complex local drug−protein interaction process.

Notably, all models achieve similarly high performance on the Human data set (AUROC > 0.97). However, this phenomenon does not necessarily imply real-world predictive accuracy. As noted in the study by Chen et al.,[26] the Human data set exists some hidden ligand bias, leading to overfitting and correct predictions based on drug features rather than pairwise interaction information. Therefore, we conducted additional experiments to predict DTIs with unknown drugs and targets.

Performance on the Test Set with Unknown Drug-Target Pairs. In drug discovery, interactions between novel drugs and targets often remain undetermined. Due to the high costs of data annotation, only a small subset of potential drug-target combinations has been experimentally validated, leading to a significant imbalance between known and unknown DTIs. Once a model has learned the features of specific drugs and

https://doi.org/10.1021/acs.jcim.4c01554
J. Chem. Inf. Model. 2024, 64, 8952−8962

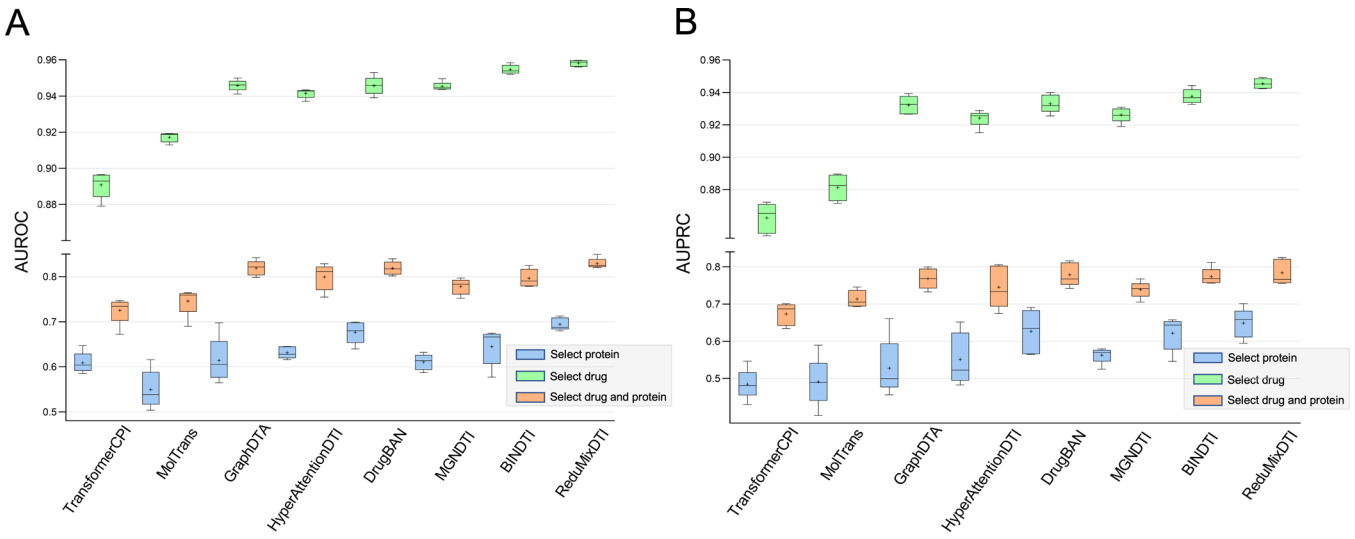


Figure 3. Performance comparison of the models on three unknown test sets of the BindingDB data set, evaluated using AUROC (A) and AUPRC (B) metrics across five random runs. Each box plot in our experiment displays five horizontal lines, representing (from top to bottom) the maximum, second-highest, median, second-lowest, and minimum values. Additionally, the plus sign indicates the mean value.

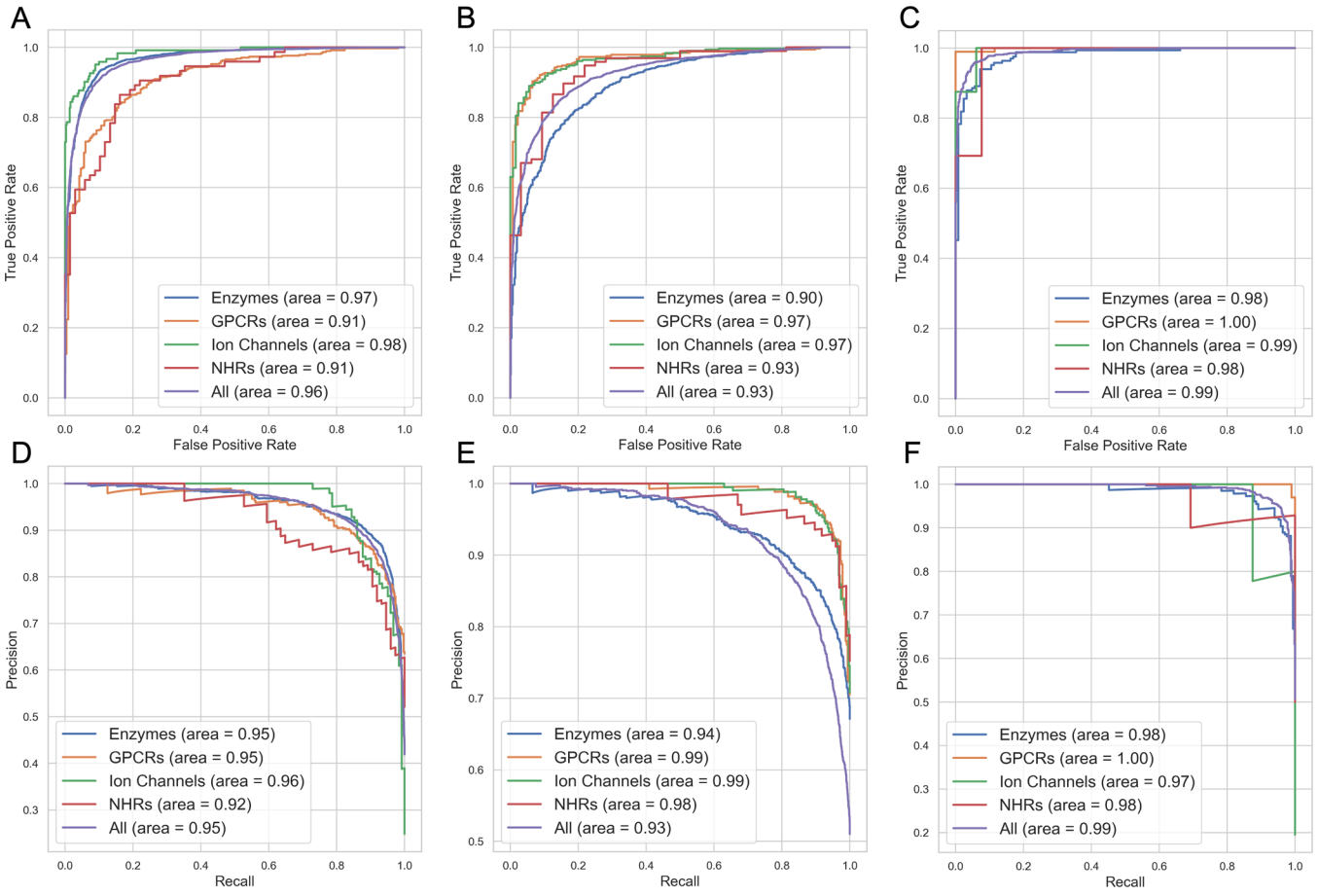


Figure 4. ReduMixDTI performance on different protein families. (A) AUROC curves on the BindingDB data set. (B) AUROC curves on the BioSNAP data set. (C) AUROC curves on the Human data set. (D) AUPRC curves on the BindingDB data set. (E) AUPRC curves on the BioSNAP data set. (F) AUPRC curves on the Human data set.

proteins as well as the interactions of a drug (or target) with other targets (or drugs), it can facilitate the identification of similar interaction patterns. However, when the model encounters unknown drugs or unknown targets or lacks information on both, the data distribution deviates from typical random splits, making prediction more challenging.

To evaluate the model's predictive ability on these challenging unknown DTIs, we conduct experiments using the BindingDB data set, which contains the largest number of validated drugs

https://doi.org/10.1021/acs.jcim.4c01554
J. Chem. Inf. Model. 2024, 64, 8952−8962

Table 5. Ablation Test Results of the CSFR Module on the BindingDB Data Set

CSFR			evaluation metrics						
baseline	ACR	SICR	AUROC (std)	AUPRC (std)	accuracy (std)	recall (std)	precision (std)	specificity (std)	F1 (std)
√			0.962 (0.002)	0.951 (0.003)	0.907 (0.008)	0.901 (0.007)	0.889 (0.007)	0.912 (0.006)	0.895 (0.004)
√	√		0.965 (0.001)	0.954 (0.002)	0.910 (0.006)	0.901 (0.008)	0.895 (0.008)	0.914 (0.007)	0.898 (0.005)
√		√	0.967 (0.003)	0.958 (0.003)	0.911 (0.010)	0.904 (0.009)	0.897 (0.007)	0.913 (0.009)	0.901 (0.006)
√	√	√	0.970 (0.003)	0.961 (0.002)	0.918 (0.006)	0.915 (0.010)	0.901 (0.012)	0.920 (0.011)	0.907 (0.008)

Table 6. Comparison Results of the MBCA Module and the Three Variants on the BindingDB Data Set

	AUROC (std)	AUPRC (std)	accuracy (std)	recall (std)	precision (std)	specificity (std)	F1 (std)
LC	0.946 (0.004)	0.938 (0.005)	0.884 (0.015)	0.873 (0.010)	0.871 (0.008)	0.898 (0.012)	0.872 (0.004)
D2PCA	0.954 (0.002)	0.945 (0.003)	0.898 (0.007)	0.885 (0.009)	0.887 (0.012)	0.910 (0.011)	0.886 (0.008)
P2DCA	0.952 (0.001)	0.943 (0.002)	0.895 (0.004)	0.882 (0.010)	0.885 (0.009)	0.909 (0.008)	0.884 (0.006)
MBCA	0.962 (0.002)	0.951 (0.003)	0.907 (0.008)	0.901 (0.007)	0.889 (0.007)	0.912 (0.006)	0.895 (0.004)

and targets among three benchmark data sets. We randomly selected 20% of the drugs, proteins, or both, designating all related DTIs as an unknown DTI set for testing. The remaining pairs are split into training and validation sets in a 7:1 ratio. Subsequently, all baseline models and our proposed model are tested on these unknown sets.

The results in Figure 3 indicate that all models experience a significant performance decline with this split strategy, particularly TransformerCPI and MolTrans. The limited number of targets relative to drugs in the data set restricts these models to a narrow set of protein features and interaction patterns, resulting in a more pronounced decline when targets and their related DTIs are unknown. Additionally, performance is intermediate when both drugs and targets are unknown, suggesting that excluding both drugs and proteins simultaneously has a balancing effect on model performance. Notably, our proposed model consistently outperforms other SOTA baselines across the three unknown scenarios, further validating its effectiveness and demonstrating its generalization potential for unknown DTI prediction.

Model Sensitivity Analysis. To further evaluate the sensitivity of the ReduMixDTI model to diverse DTI data sets, we conduct experiments to assess its performance across various protein families. Based on the research of Bai et al.,[17] we select four major protein families: enzymes, G protein-coupled receptors (GPCRs), ion channels, and nuclear hormone receptors (NHRs). We randomly sample data from three benchmark data sets as the test set and utilize the GtoPdb database[46] to map the proteins in this set to the four families, forming corresponding DTI data sets. Figure 4 demonstrates that the performance of ReduMixDTI (measured by AUROC and AUPRC) remains consistent across these protein family data sets, further confirming the model's applicability and robustness across various data sets.

Ablation Studies. To thoroughly evaluate the influence of the CSFR and MBCA modules, we conduct detailed ablation studies on the BindingDB data set over five random runs. The ReduMixDTI model without the CSFR module serves as the baseline. By comparing the performance of methods incorporating different components of the CSFR module, we assess the contributions of the ACR and SICR units with constituent units of each method indicated by checkmarks in Table 5.

Additionally, we develop three variants to evaluate the MBCA module's effect: linear concatenation (LC), drug-to-protein one-sided cross-attention (D2PCA), and protein-to-drug one-sided cross-attention (P2DCA). In the LC variant, drug and protein vectors are concatenated after max-pooling operations. The D2PCA variant captures joint representation by feeding a drug vector and a protein subsequence matrix into multihead cross attention within the Transformer decoder.[23] Conversely, the P2DCA variant uses a drug substructure matrix and a protein vector for joint representation. We replace the MBCA module in the baseline with these three variants to assess the model performance under different pair representation computation methods, with results summarized in Table 6. The detailed analysis of the CSFR and MBCA modules' effectiveness follows in subsequent sections.

Effectiveness of the CSFR. Given that powerful feature encoders may extract global features irrelevant to DTI prediction, we propose the CSFR module to reduce the feature redundancy. Compared to the baseline model, adding the ACR unit improves performance, demonstrating the benefit of incorporating directional and multiscaled information in the feature map. By adjusting channel importance based on this information and integrating multiscaled features, the model's focus on key features and its expressive capacity are enhanced. Additionally, the improvement observed when adding the SICR unit to the baseline suggests that spatial content segmentation and cross-reorganization reinforce useful feature information. Using the complete CSFR module leads to further improvements in evaluation metrics, indicating that even channel-refined features exhibit spatial redundancy. Combining channel attention with spatial splitting enables the model to dynamically adjust and capture the most essential features for the DTI prediction.

Effectiveness of the MBCA. We propose a novel hybrid attention module named MBCA to model pairwise local interactions between drugs and proteins. Leveraging linear transformations, MBCA adaptively integrates diverse interaction possibilities captured by multihead attention for comprehensive representation of complex interactions. As shown in Table 6, the metrics for directly concatenating drug and protein target features are lower than those for the two variants employing single-sided cross-attention. These results highlight the importance of capturing drug−protein interactions for accurate DTI prediction with single-sided attention on either drugs or proteins specifically capturing partial binding information. In contrast, the MBCA module, which integrates attentional insights from both drugs and proteins to better represent the binding regions of the complexes, demonstrates superior performance in fine-grained modeling of drug−protein local interactions.

https://doi.org/10.1021/acs.jcim.4c01554
J. Chem. Inf. Model. 2024, 64, 8952−8962

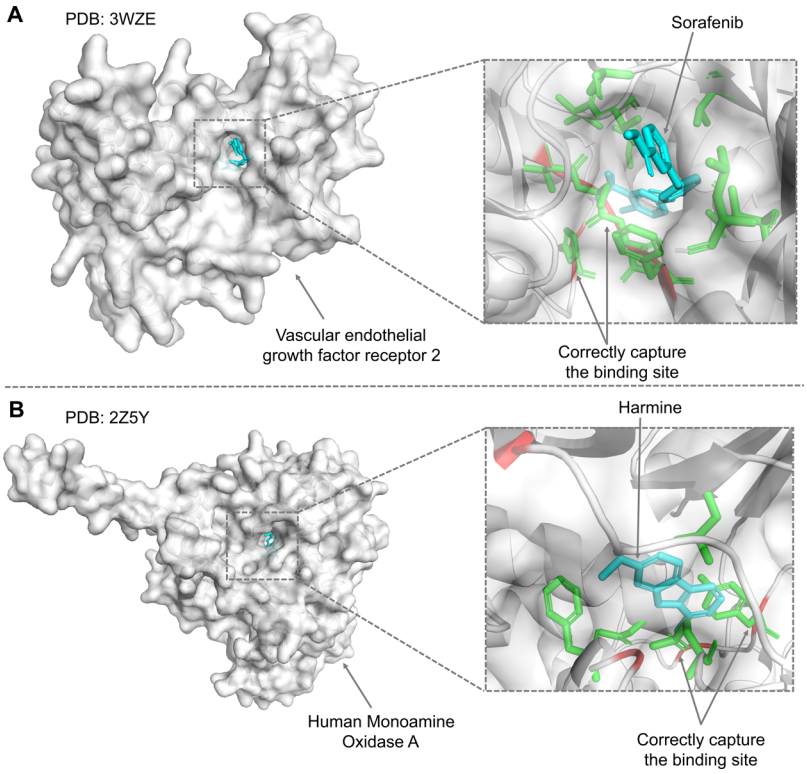


Figure 5. Visualization of predicted protein attention weights and binding pockets for interpretability. The 3D representations of ligand-protein binding pockets are shown in (A) (PDB ID: 3WZE) and (B) (PDB ID: 2Z5Y). In the enlarged binding pockets, ligands are shown in cyan, and proteins are in white. Additionally, the actual binding sites are highlighted in green, and the predicted sites of our model are shown in red.

Interpretation Studies. Finally, we conducted experiments on two cases to obtain protein attention weights representing the contribution of each potential substructure to the DTI prediction. Figure 5A shows the cocrystallized complex of Vascular Endothelial Growth Factor Receptor 2 with Sorafenib (PDB ID: 3WZE[47]), while Figure 5B presents the complex of Human Monoamine Oxidase A with Harmine (PDB ID: 2Z5Y[48]).

The drug SMILES and protein sequence are input into the ReduMixDTI model. After obtaining the protein attention matrix $\mathbf{A}_p$, we apply average pooling to produce an attention vector $a_p \in \mathbb{R}^N$, which is mapped onto the known 3D protein structure. The drug-target complexes are visualized in 3D, with proteins in white and drugs in cyan. Additionally, residues with high attention weights are highlighted in red, while true binding sites are marked in green. In the case of 3WZE, two predicted regions align with the true binding sites. Similarly, in 2Z5Y, two predicted regions match true binding sites with one region close to the true binding sites. These visualization results suggest that our MBCA method has the potential to recognize DTI combinations. However, we also observe that predicted correct sites constitute only a small part of the true sites and that some regions deviate from known binding sites, which may indicate that the attention mechanism does not fully capture actual binding sites or may reveal unexplored sites of local interactions.

CONCLUSIONS

In this work, we propose ReduMixDTI, a novel DTI prediction model that minimizes global feature redundancy and emphasizes critical local interaction features essential for the DTI prediction. To this end, we introduce a redundancy removal module that reconstructs features based on channel interdependencies and spatial feature reorganization. Additionally, the specialized attention module focuses on key interaction features, significantly improving the model's predictive ability. Compared to other SOTA DTI models across three benchmark data sets, ReduMixDTI consistently demonstrates superior performance under both random splits and splits-based real-world scenarios. Furthermore, by visualizing attention weights on 3D drug−protein complexes, our model identifies potential binding sites and provides biological insights into interaction processes.

Although our work has demonstrated an effective performance in DTI classification, several limitations should be addressed in future research. First, our method predicts DTIs based on 1D protein sequences and 2D molecular graphs without considering 3D structures of drugs and proteins. Given the significant advancements made by DeepMind's AlphaFold3[49] in predicting protein 3D structures, integrating high-quality structural data could enhance both the performance and interpretability of DTI prediction models. Additionally, incorporating related information, such as drug−drug interactions, drug-disease associations, protein−protein interactions, and protein-disease associations, could provide valuable knowledge outside the field, enriching the model's understanding of drugs, proteins, and DTI processes. Second, the predicted binding sites in our model do not fully represent the actual binding sites. Future research could incorporate binding site labels during training to improve localization accuracy and provide deeper insights into DTIs. Third, enhancing model generalization for unknown DTIs remains challenging. Exploring semisupervised and unsupervised methods could facilitate the extraction of more valuable information from unknown DTIs.

https://doi.org/10.1021/acs.jcim.4c01554
J. Chem. Inf. Model. 2024, 64, 8952−8962

Furthermore, the proposed idea for DTI prediction is general and can be extended to other interaction prediction tasks such as small molecule-miRNA association (MMA) prediction. The high noise levels and limited MMA data availability present challenges for current models.[50−52] In contrast, DTI data are more abundant. A promising direction for future research could involve leveraging transfer learning to fine-tune models trained on DTI data sets for these tasks.

■ ASSOCIATED CONTENT

Data Availability Statement

The data and codes of our model are available at https://github.com/mql430/ReduMixDTI.

■ AUTHOR INFORMATION

Corresponding Authors

Hongqi Li − *Department of Geriatrics, The First Affiliated Hospital of USTC, University of Science and Technology of China, Hefei 230026 Anhui, China*; Email: redflag_li@163.com

Ji Liu − *National Engineering Laboratory for Brain-inspired Intelligence Technology and Application, School of Information Science and Technology and MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China, Hefei 230026 Anhui, China; Center for Advanced Interdisciplinary Science and Biomedicine of IHM, Division of Life Sciences and Medicine, University of Science and Technology of China MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China, Hefei 230026 Anhui, China; Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230026 Anhui, China*; Email: lj1257@ustc.edu.cn

Authors

Mingqing Liu − *National Engineering Laboratory for Brain-inspired Intelligence Technology and Application, School of Information Science and Technology, University of Science and Technology of China, Hefei 230026 Anhui, China; Center for Advanced Interdisciplinary Science and Biomedicine of IHM, Division of Life Sciences and Medicine, University of Science and Technology of China MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China, Hefei 230026 Anhui, China*; orcid.org/0009-0007-7314-6472

Xuechun Meng − *National Engineering Laboratory for Brain-inspired Intelligence Technology and Application, School of Information Science and Technology, University of Science and Technology of China, Hefei 230026 Anhui, China; Center for Advanced Interdisciplinary Science and Biomedicine of IHM, Division of Life Sciences and Medicine, University of Science and Technology of China MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China, Hefei 230026 Anhui, China*

Yiyang Mao − *National Engineering Laboratory for Brain-inspired Intelligence Technology and Application, School of Information Science and Technology, University of Science and Technology of China, Hefei 230026 Anhui, China; Center for Advanced Interdisciplinary Science and Biomedicine of IHM, Division of Life Sciences and Medicine, University of Science and Technology of China MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China, Hefei 230026 Anhui, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.4c01554

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (91957112 to J.L.), the National Natural Science Foundation of China (31970950 to J.L.), and the Natural Science Foundation of Anhui Province (2308085MH239 to H.L.). We acknowledge the support of the GPU cluster built by the MCC Lab of Information Science and Technology Institution, USTC.

■ REFERENCES

(1) Tabei, Y.; Pauwels, E.; Stoven, V.; Takemoto, K.; Yamanishi, Y. Identification of Chemogenomic Features from Drug−Target Interaction Networks Using Interpretable Classifiers. *Bioinformatics* **2012**, *28*, i487−i494.

(2) Sachdev, K.; Gupta, M. K. A Comprehensive Review of Feature Based Methods for Drug Target Interaction Prediction. *Journal of Biomedical Informatics* **2019**, *93*, No. 103159.

(3) Pushpakom, S.; Iorio, F.; Eyers, P. A.; Escott, K. J.; Hopper, S.; Wells, A.; Doig, A.; Guilliams, T.; Latimer, J.; McNamee, C.; Norris, A.; Sanseau, P.; Cavalla, D.; Pirmohamed, M. Drug Repurposing: Progress, Challenges and Recommendations. *Nat. Rev. Drug Discovery* **2019**, *18*, 41−58.

(4) Jarada, T. N.; Rokne, J. G.; Alhajj, R. A Review of Computational Drug Repositioning: Strategies, Approaches, Opportunities, Challenges, and Directions. *J. Cheminform.* **2020**, *12*, 1−23.

(5) Li, J.; Zheng, S.; Chen, B.; Butte, A. J.; Swamidass, S. J.; Lu, Z. A Survey of Current Trends in Computational Drug Repositioning. *Briefings in bioinformatics* **2016**, *17*, 2−12.

(6) Wishart, D. S.; et al. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074−D1082.

(7) DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The Price of Innovation: New Estimates of Drug Development Costs. *Journal of Health Economics* **2003**, *22*, 151−185.

(8) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge. *Nat. Rev. Drug Discovery* **2010**, *9*, 203−214.

(9) Cichonska, A.; Ravikumar, B.; Parri, E.; Timonen, S.; Pahikkala, T.; Airola, A.; Wennerberg, K.; Rousu, J.; Aittokallio, T. Computational-Experimental Approach to Drug-Target Interaction Mapping: A Case Study on Kinase Inhibitors. *PLOS Computational Biology* **2017**, *13*, No. e1005678.

(10) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of computational chemistry* **2009**, *30*, 2785−2791.

(11) Maia, E. H. B.; Assis, L. C.; de Oliveira, T. A.; da Silva, A. M.; Taranto, A. G. Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Front. Chem.* **2020**, *8*, 343.

(12) Allenspach, S.; Hiss, J. A.; Schneider, G. Neural Multi-Task Learning in Drug Design. *Nature Machine Intelligence* **2024**, *6*, 124−137.

(13) Lim, S.; Lu, Y.; Cho, C. Y.; Sung, I.; Kim, J.; Kim, Y.; Park, S.; Kim, S. A Review on Compound-Protein Interaction Prediction Methods: Data, Format, Representation and Model. *Computational and Structural Biotechnology Journal* **2021**, *19*, 1541−1556.

(14) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.

(15) Peng, J.; Wang, Y.; Guan, J.; Li, J.; Han, R.; Hao, J.; Wei, Z.; Shang, X. An End-to-End Heterogeneous Graph Representation Learning-Based Framework for Drug−Target Interaction Prediction. *Briefings Bioinf.* **2021**, *22*, No. bbaa430.
(16) Zhao, Q.; Zhao, H.; Zheng, K.; Wang, J. HyperAttentionDTI: Improving Drug−Protein Interaction Prediction by Sequence-Based Deep Learning with Attention Mechanism. *Bioinformatics* **2022**, *38*, 655−662.
(17) Bai, P.; Miljković, F.; John, B.; Lu, H. Interpretable Bilinear Attention Network with Domain Adaptation Improves Drug−Target Prediction. *Nature Machine Intelligence* **2023**, *5*, 126−136.
(18) Öztürk, H.; Özgür, A.; Ozkirimli, E. DeepDTA: Deep Drug−Target Binding Affinity Prediction. *Bioinformatics* **2018**, *34*, i821−i829.
(19) Tsubaki, M.; Tomii, K.; Sese, J. Compound−Protein Interaction Prediction with End-to-End Learning of Neural Networks for Graphs and Sequences. *Bioinformatics* **2019**, *35*, 309−318.
(20) Abbasi, K.; Razzaghi, P.; Poso, A.; Amanlou, M.; Ghasemi, J. B.; Masoudi-Nejad, A. DeepCDA: Deep Cross-Domain Compound−Protein Affinity Prediction through LSTM and Convolutional Neural Networks. *Bioinformatics* **2020**, *36*, 4633−4642.
(21) Huang, K.; Xiao, C.; Glass, L. M.; Sun, J. MolTrans: Molecular Interaction Transformer for Drug−Target Interaction Prediction. *Bioinformatics* **2021**, *37*, 830−836.
(22) Zhang, L.; Wang, C.-C.; Chen, X. Predicting Drug-Target Binding Affinity through Molecule Representation Block Based on Multi-Head Attention and Skip Connection. *Briefings Bioinf.* **2022**, *23*, bbac468.
(23) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I.; *Attention Is All You Need*. In Advances in Neural Information Processing Systems, 2017
(24) Zhang, L.; Wang, C.-C.; Zhang, Y.; Chen, X. GPCNDTA: Prediction of Drug-Target Binding Affinity through Cross-Attention Networks Augmented with Graph Features and Pharmacophores. *Computers in Biology and Medicine* **2023**, *166*, No. 107512.
(25) Zhu, Z.; Yao, Z.; Qi, G.; Mazur, N.; Yang, P.; Cong, B. Associative Learning Mechanism for Drug-Target Interaction Prediction. *CAAI Transactions on Intelligence Technology* **2023**, *8*, 1558−1577.
(26) Chen, L.; Tan, X.; Wang, D.; Zhong, F.; Liu, X.; Yang, T.; Luo, X.; Chen, K.; Jiang, H.; Zheng, M. TransformerCPI: Improving Compound−Protein Interaction Prediction by Sequence-Based Deep Learning with Self-Attention Mechanism and Label Reversal Experiments. *Bioinformatics* **2020**, *36*, 4406−4414.
(27) Kim, J.-H.; Jun, J.; Zhang, B.-T.; Bilinear Attention Networks. In *Advances in Neural Information Processing Systems*, 2018.
(28) Zhu, Z.; Yao, Z.; Zheng, X.; Qi, G.; Li, Y.; Mazur, N.; Gao, X.; Gong, Y.; Cong, B. Drug−Target Affinity Prediction Method Based on Multi-Scale Information Interaction and Graph Optimization. *Computers in Biology and Medicine* **2023**, *167*, No. 107621.
(29) Zhu, Z.; Zheng, X.; Qi, G.; Gong, Y.; Li, Y.; Mazur, N.; Cong, B.; Gao, X. Drug−Target Binding Affinity Prediction Model Based on Multi-Scale Diffusion and Interactive Learning. *Expert Systems with Applications* **2024**, *255*, No. 124647.
(30) Ferreira De Freitas, R.; Schapira, M. A Systematic Analysis of Atomic Protein−Ligand Interactions in the PDB. *MedChemComm* **2017**, *8*, 1970−1981.
(31) Pichler, W. J. The Important Role of Non-Covalent Drug-Protein Interactions in Drug Hypersensitivity Reactions. *Allergy* **2022**, *77*, 404−415.
(32) Zuo, Y.; Wu, X.; Ge, F.; Yan, H.; Fei, S.; Liang, J.; Deng, Z. Research Progress on Drug-Target Interactions in the Last Five Years. *Anal. Biochem.* **2024**, *697*, No. 115691.
(33) Liu, H.; Sun, J.; Guan, J.; Zheng, J.; Zhou, S. Improving Compound−Protein Interaction Prediction by Building up Highly Credible Negative Samples. *Bioinformatics* **2015**, *31*, i221−i229.
(34) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: A Knowledgebase for Drugs, Drug Actions and Drug Targets. *Nucleic acids research* **2008**, *36*, D901−D906.
(35) Gao, K. Y.; Fokoue, A.; Luo, H.; Iyengar, A.; Dey, S.; Zhang, P. Interpretable Drug Target Prediction Using Deep Neural Representation. *Electron. Proc. IJCAI* **2018**, *2018*, 3371−3377.
(36) Bai, P.; Miljković, F.; Ge, Y.; Greene, N.; John, B.; Lu, H.*Hierarchical Clustering Split for Low-Bias Evaluation of Drug-Target Interaction Prediction*. In IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021; pp. 641−644.
(37) Zitnik, M.; Sosic, R.; Leskovec, J. BioSNAP Datasets: Stanford Biomedical Network Dataset Collection. http://snap.stanford.edu/biodata 2018
(38) Li, M.; Zhou, J.; Hu, J.; Fan, W.; Zhang, Y.; Gu, Y.; Karypis, G. DGL-LifeSci: An Open-Source Toolkit for Deep Learning on Graphs in Life Science. *ACS Omega* **2021**, *6*, 27233−27238.
(39) Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z.; *Efficient Multi-Scale Attention Module with Cross-Spatial Learning*. In ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023; pp 1−5.
(40) Li, J.; Wen, Y.; He, L.; *SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023; pp 6153−6162.
(41) Nguyen, T.; Le, H.; Quinn, T. P.; Nguyen, T.; Le, T. D.; Venkatesh, S. GraphDTA: Predicting Drug−Target Binding Affinity with Graph Neural Networks. *Bioinformatics* **2021**, *37*, 1140−1147.
(42) Peng, L.; Liu, X.; Chen, M.; Liao, W.; Mao, J.; Zhou, L. MGNDTI: A Drug-Target Interaction Prediction Framework Based on Multimodal Representation Learning and the Gating Mechanism. *J. Chem. Inf. Model.* **2024**, *64*, 6684−6698.
(43) Sun, Y.; Dong, L.; Huang, S.; Ma, S.; Xia, Y.; Xue, J.; Wang, J.; Wei, F. *Retentive Network: A Successor to Transformer for Large Language Models*, 2023.
(44) Peng, L.; Liu, X.; Yang, L.; Liu, L.; Bai, Z.; Chen, M.; Lu, X.; Nie, L. BINDTI: A Bi-Directional Intention Network for Drug-Target Interaction Identification Based on Attention Mechanisms. *IEEE Journal of Biomedical and Health Informatics* **2024**, 1−11.
(45) Garnelo, M.; Czarnecki, W. M. *Exploring the Space of Key-Value-Query Models with Intention*, 2023.
(46) Harding, S. D.; Armstrong, J. F.; Faccenda, E.; Southan, C.; Alexander, S. P. H.; Davenport, A. P.; Spedding, M.; Davies, J. A. The IUPHAR/BPS Guide to PHARMACOLOGY in 2024. *Nucleic Acids Res.* **2024**, *52*, D1438−D1449.
(47) Okamoto, K.; Ikemori-Kawada, M.; Jestel, A.; von König, K.; Funahashi, Y.; Matsushima, T.; Tsuruoka, A.; Inoue, A.; Matsui, J. Distinct Binding Mode of Multikinase Inhibitor Lenvatinib Revealed by Biochemical Characterization. *ACS Med. Chem. Lett.* **2015**, *6*, 89−94.
(48) Son, S.-Y.; Ma, J.; Kondou, Y.; Yoshimura, M.; Yamashita, E.; Tsukihara, T. Structure of Human Monoamine Oxidase A at 2.2-Å Resolution: The Control of Opening the Entry for Substrates/Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 5739−5744.
(49) Abramson, J.; et al. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630*, 493−500.
(50) Chen, X.; Guan, N.-N.; Sun, Y.-Z.; Li, J.-Q.; Qu, J. MicroRNA-small Molecule Association Identification: From Experimental Results to Computational Models. *Briefings Bioinf.* **2018**, *21*, 47−61.
(51) Zhou, Z.; Zhuo, L.; Fu, X.; Lv, J.; Zou, Q.; Qi, R. Joint Masking and Self-Supervised Strategies for Inferring Small Molecule-miRNA Associations. *Molecular Therapy - Nucleic Acids* **2024**, *35*, No. 102103.
(52) Zhou, Z.; Du, Z.; Jiang, X.; Zhuo, L.; Xu, Y.; Fu, X.; Liu, M.; Zou, Q. GAM-MDR: Probing miRNA-drug Resistance Using a Graph Autoencoder Based on Random Path Masking. *Briefings in Functional Genomics* **2024**, *23*, 475−483.

https://doi.org/10.1021/acs.jcim.4c01554
J. Chem. Inf. Model. 2024, 64, 8952−8962