



Research article

Behavior-aware contextual representation learning for drug-target interaction prediction

Yimin Yin^{1,2}, Jianying Qu^{1,2}, Yingying Li³, Jing Zhang⁴, Bin Yan^{1,2}, Wanxia Deng^{5,*} and Jinghua Zhang^{6,*}

¹ School of Mathematics and Statistics, Hunan First Normal University, Changsha 410205, China

² Hunan Provincial University Key Laboratory for Big Data Analysis and Application, Hunan First Normal University, Changsha 410205, China

³ College of Science, National University of Defense Technology, Changsha 410073, China

⁴ Key Laboratory of Computing and Stochastic Mathematics, Hunan Normal University, Changsha 410081, China

⁵ College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China

⁶ College of Computer Science and Software Engineering, Hohai University, Nanjing 211100, China

* **Correspondence:** Email: wanxiadeng@163.com, zhangjingh@foxmail.com.

Abstract: Accurate prediction of drug-target interactions (DTIs) is a fundamental task in computational drug discovery, as it can effectively reduce the cost and time required for experimental screening. However, although most existing methods consider both sides of an interaction, they still focus mainly on the intrinsic features of drugs and proteins and are therefore insufficient to capture higher-order behavioral context and complementary interaction patterns in DTIs. To address this issue, we propose a novel DTI prediction framework that integrates drug molecular structure, protein sequence information, and behavior-aware contextual representations within a unified architecture. Specifically, drug compounds are encoded by a graph neural network (GNN) to capture molecular topology, protein sequences are modeled to learn semantic features, and a protein-centric behavioral sequence is introduced to characterize pharmacological interaction patterns. In addition, an attention-based fusion mechanism is introduced to model dependencies within the behavioral sequence and selectively aggregate informative behavioral contextual signals. Extensive experiments on three benchmark datasets demonstrate that the proposed method consistently outperforms representative baseline models. These results indicate that the proposed framework provides an effective computational solution for DTI prediction.

Keywords: drug-target interaction prediction; graph neural networks; behavioral sequence; attention mechanism

1. Introduction

Drug discovery is widely recognized as an expensive, time-consuming, and high-risk process. Bringing a single new drug from initial discovery to clinical approval typically requires more than a decade of development and an investment exceeding one billion US dollars, while the failure rate across clinical trial phases remains above 90% [1–3]. One of the major bottlenecks in this pipeline is the identification of drug-target interactions (DTIs), which aims to determine whether a small-molecule compound can selectively bind to a specific protein target and thereby modulate its biological activity. Traditionally, this process has relied heavily on high-throughput screening and biochemical binding assays. Although experimentally rigorous, these approaches are limited by substantial reagent costs, relatively low throughput relative to the vast chemical space, and the practical impossibility of exhaustively testing all candidate compound-protein pairs [4, 5]. Meanwhile, the number of known bioactive compounds and sequenced proteins has grown rapidly in recent years, with databases such as ChEMBL [6] and UniProt [7, 8] now cataloging millions of entries. As a result, the combinatorial scale of potential interactions makes purely wet-lab-based approaches fundamentally inadequate. Therefore, developing accurate and scalable computational methods for DTI prediction has become an indispensable component of modern drug discovery, as such methods can prioritize the most promising candidates before costly experimental validation [9].

Early computational approaches to DTI prediction can be broadly categorized into structure-based and similarity-based methods. Structure-based methods, most notably molecular docking, estimate binding affinity by simulating the physical interaction between a ligand and a protein binding site, and they have achieved considerable success in well-characterized target families [10]. However, these methods are fundamentally constrained by their reliance on high-resolution three-dimensional protein structures, which remain unavailable for a large proportion of therapeutically relevant targets. In addition, their substantial computational cost makes large-scale virtual screening impractical. Similarity-based methods, by contrast, operate under the assumption that chemically similar compounds tend to interact with similar protein targets, and they exploit known interaction profiles to infer novel associations [11]. Although conceptually straightforward, these approaches suffer from limited generalizability because they are inherently unable to predict interactions involving compounds or proteins that are structurally dissimilar to those already present in existing databases. To address these shortcomings, machine learning methods, including matrix factorization [12], random forests [13], and support vector machines [14–16], were subsequently introduced. These methods leverage handcrafted molecular fingerprints and physicochemical descriptors to model DTI patterns. Nevertheless, their predictive performance remains heavily dependent on the quality and coverage of manually engineered features, and their ability to capture the complex nonlinear relationships underlying molecular recognition is fundamentally limited. In recent years, deep learning has emerged as a transformative paradigm for DTI prediction, offering the ability to automatically learn hierarchical representations directly from raw molecular data and substantially reducing reliance on domain-specific feature engineering.

Building on the strong representational power of deep learning, a substantial body of work has explored neural architectures tailored to the structural properties of drugs and proteins. On the drug side, graph neural networks (GNNs) have become a dominant paradigm for molecular representation learning, as they naturally encode atoms as nodes and chemical bonds as edges, thereby enabling models to capture both local chemical environments and global topological properties of molecular graphs [17, 18].

Representative methods such as graph convolutional networks (GCNs) [19], graph attention networks (GATs) [20], and their domain-adapted variants have demonstrated strong performance in molecular property prediction tasks and have been widely adopted in DTI frameworks [21]. On the protein side, convolutional neural networks (CNNs) [22] applied to integer-encoded amino acid sequences have proved effective in extracting local motif patterns from raw sequence data without requiring explicit structural annotations. More recently, transformer-based architectures and pretrained protein language models such as evolutionary scale modeling (ESM) have further improved protein representation learning by capturing long-range residue dependencies through self-attention mechanisms. By integrating these two streams of representation, a number of end-to-end DTI models have been proposed that jointly encode drug and protein features and then fuse them through concatenation, bilinear interaction, or cross-attention modules to generate interaction predictions. Despite these advances, a fundamental limitation shared by the vast majority of existing approaches is that they model each drug–protein pair in isolation, relying primarily on the intrinsic physicochemical properties of the individual molecule and the target sequence. This pairwise paradigm fails to exploit the rich relational context embedded in the broader interaction network and, in particular, overlooks the behavioral profile of a protein, namely the set of known binding partners with which it interacts and the corresponding interaction strengths. Such collective interaction patterns are highly informative for characterizing a protein’s binding preferences and selectivity, yet they remain systematically underutilized by current deep learning methods, leaving clear room for further improvement.

To address the aforementioned limitation, we propose a novel deep learning framework for DTI prediction that explicitly incorporates the historical interaction behavior of proteins as a complementary source of contextual information. Our approach is inspired by collaborative filtering in recommender systems, where a user’s preference profile, constructed from past interactions with items, provides strong predictive signals for future behavior. Analogously, we argue that a protein does not exist in isolation within the biochemical interaction landscape. Rather, the set of compounds to which it has previously been observed to bind, together with the corresponding interaction strengths, constitutes a rich and informative behavioral fingerprint that reflects its underlying binding preferences and selectivity in a data-driven manner. From this perspective, DTI prediction can be reframed from a pairwise matching problem into a context-aware reasoning task, in which the relational history of the target protein serves as a form of soft prior knowledge that guides the model toward more informed and generalizable interaction predictions. Building on this motivation, we design a framework that jointly encodes drug molecular structure, protein sequence, and protein behavioral context, and integrates them through an attention-based fusion mechanism that adaptively weights historical interaction partners according to their observed binding strengths.

The main contributions of this work are summarized as follows. First, we introduce a collaborative-filtering-inspired behavioral modeling strategy into the DTI prediction framework, which, to our knowledge, has not been systematically explored in prior DTI studies. Second, we design an attention-based behavior fusion module that incorporates interaction strength as an explicit weighting signal, enabling adaptive aggregation of protein behavioral context. Finally, extensive experiments conducted on benchmark datasets demonstrate that our proposed model achieves competitive performance against state-of-the-art baselines across multiple evaluation metrics.

2. Related work

Deep learning, driven by large-scale data, has revolutionized many fields [23–26] and has also made significant progress in DTI prediction. Deep learning models can automatically learn rich feature representations of drugs and proteins from data, achieving performance that far surpasses that of traditional DTI methods. For drug and protein representation learning, CNN-based and attention-based methods have been widely adopted. Deep Drug-target affinity (DeepDTA) [27] employs two independent one-dimensional CNN (1D-CNN) modules to process drug simplified molecular input line entry system (SMILES) strings and protein amino acid sequences, respectively, extracting local patterns that are subsequently concatenated and fed into a fully connected layer for prediction. Deep convolutional DTI (DeepConv-DTI) [28] uses a CNN to extract features from protein sequences and a fully connected network to process drug features for DTI prediction. Specifically, it employs a CNN followed by a global max-pooling layer to capture local patterns of varying lengths within protein sequences, while a fully connected layer is used to process drug fingerprints. Deep drug-target binding affinity (DeepDTAF) [29] proposes a deep learning framework that integrates local and global features. The model consists of three independent modules: a global protein sequence module, a local binding pocket module, and a ligand module. The protein and pocket features incorporate secondary structure elements and physicochemical properties. In addition, the model introduces dilated convolutions with varying dilation rates in the protein and ligand modules to capture multiscale long-range interaction information within the sequences. Molecular interaction transformer (MolTrans) [30] develops a knowledge-inspired algorithm for substructure pattern mining and designs a biomimetic interactive modeling module based on an enhanced transformer encoder to capture semantic relationships between substructures from large amounts of unlabeled biomedical data for DTI prediction. Hyper attention DTI (HyperAttentionDTI) [31] takes drug SMILES strings and protein amino acid sequences as input, extracts a feature matrix using stacked 1D-CNNs, and then employs a custom-designed HyperAttention module to simultaneously model complex non-covalent interactions between drugs and proteins in both spatial and channel dimensions, assigning an independent attention vector to each atom–amino acid pair. Drug bilinear attention network (DrugBAN) [32] proposes an interpretable bilinear attention network that combines domain adaptation techniques to model local interactions between drug molecules and target protein sequences.

Because GNNs can naturally represent drug molecules as graphs composed of atomic nodes and chemical bonds, GNN-based methods can directly encode molecular topology and have recently achieved remarkable progress in biologically relevant association-prediction tasks [33]. Graph early fusion affinity (GEFA) [34] integrates drug molecular graphs into the representation learning of protein graphs through an attention mechanism. Specifically, it uses GCNs to process the drug molecular graph and the protein graph constructed from contact maps separately, and then concatenates the fused outputs for prediction through a fully connected layer. Drug bilinear attention network (DrugBAN) [32] uses GCNs and CNNs to process drug molecular graphs and 1D protein sequences, respectively, and proposes a deep bilinear attention network that combines domain-adaptive techniques to explicitly learn pairwise local interactions between drugs and targets. Bi-directional intention network for (BINDTI) [35] proposes a hybrid model for protein feature encoding that integrates convolutional and self-attention mechanisms. It uses a GCN to extract drug graph features and employs a bidirectional attention network to fuse drug and protein features through multi-head attention. General, powerful, and scalable DTI

(GPS-DTI) [36] combines a graph isomorphism network with edge features (GINE) and a multi-head attention mechanism (MHAM); the former aggregates local structural information from neighboring nodes and edges, whereas the latter models global dependencies between atoms. Protein features are extracted using the pretrained language model ESM-2 and further refined via a CNN. Finally, a cross-attention module is introduced to dynamically fuse drug and protein features. Recent studies have further explored ensemble learning and global–local representation strategies for DTI prediction. Efficient deep model ensemble Framework for DTI (EADTI) [37] introduces an efficient deep model ensemble framework with feature adaptation and clustering-enhanced parameter fine-tuning to improve prediction performance and interpretability. A global–local framework [38] integrates local and global features of drugs and proteins, using graph-based similarity measures and a diffusion mechanism to capture global dependencies.

Although DTI methods based on GNNs have achieved substantial progress, relying solely on graph-based representations still entails several inherent limitations. First, the dependence of GNNs on known association information constrains their expressive capacity. Existing GNN-based methods rely heavily on network topology, yet biomedical networks themselves are often incomplete and noisy. Moreover, these graph structures depend strongly on known associations from both the drug and target sides; when such information is missing, model performance can deteriorate markedly. To address these limitations, bidirectional behavior learning for drug repositioning (BiBLDR) [39] reformulates drug repositioning as a behavioral sequence learning task. By constructing behavioral sequences from both drug-side and disease-side data and introducing a two-stage learning strategy, it captures the interactive relationships between drug and disease features within these bidirectional sequences more precisely. Motivated by this idea, we introduce behavioral sequence modeling into the DTI prediction task and propose a drug–target interaction prediction framework that incorporates the behavioral sequence of target proteins. Specifically, our method constructs a protein-centric behavioral sequence for each target protein to characterize its pharmacological interaction patterns. Unlike BiBLDR, which constructs bidirectional drug–disease behavioral sequences and relies on a two-stage learning strategy, our framework is designed specifically for DTI prediction and models the drugs that have historically interacted with each target protein as a protein-centric behavioral context. Rather than representing behavioral relations through predefined similarity information, the query drug and behavioral drugs are encoded by a shared GCN, while interaction-strength weighting and multi-head attention are used to adaptively aggregate structurally informative historical interactions into the target representation.

3. Methodology

3.1. Overall framework

As illustrated in Figure 1, given a drug–protein pair, the model takes three types of input: the SMILES string of the candidate drug compound, the amino acid sequence of the target protein, and the behavioral sequence of the target protein constructed from the dataset. The behavioral sequence of a protein is defined as the ordered collection of drug compounds with which the protein has been observed to interact in the dataset, along with their corresponding experimentally measured interaction labels, which serve as a proxy for binding strength.

The framework proceeds through four sequential stages. In the first stage, the candidate drug is represented as a molecular graph and processed by the drug encoder, which is implemented as a GCN,

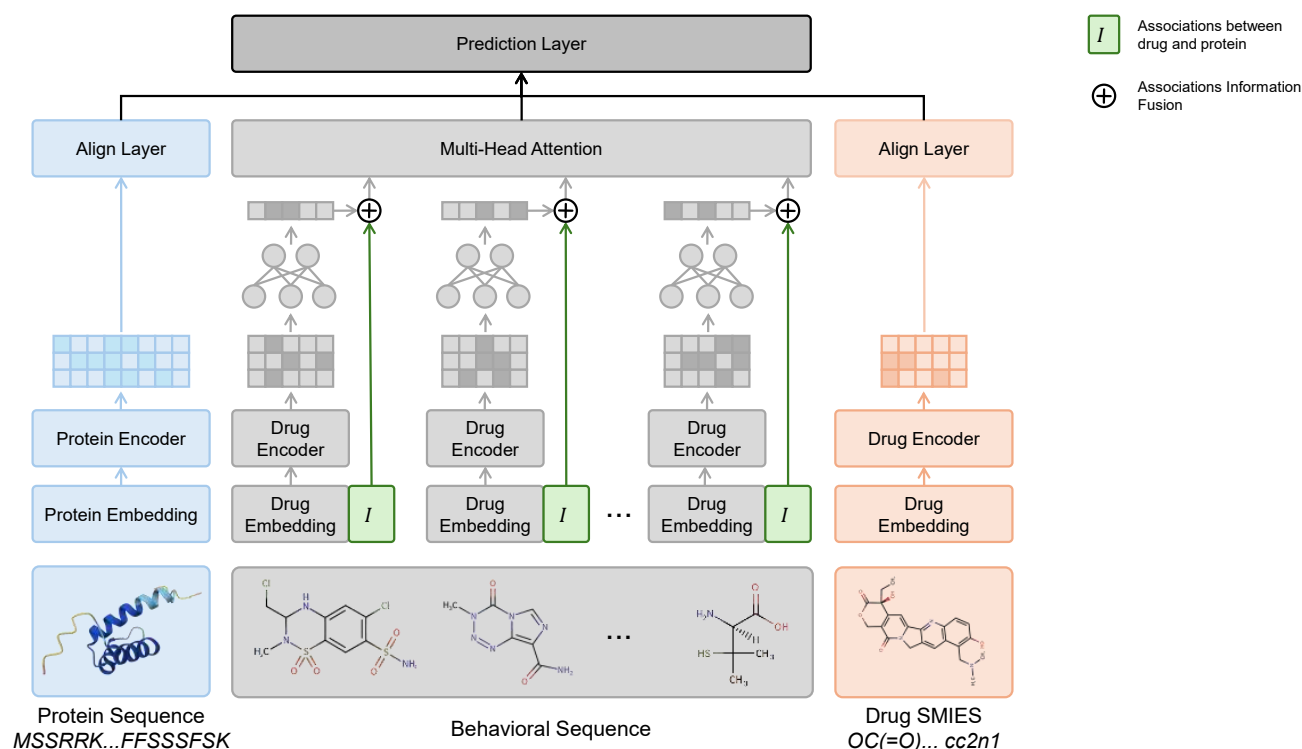


Figure 1. Schematic overview of the proposed framework for drug-target interaction prediction, integrating drug molecular encoding, protein sequence encoding, behavior-aware contextual modeling, and attention-based feature fusion within a unified architecture.

to produce a graph-based molecular representation. Concurrently, the target protein is processed from its amino acid sequence by the protein encoder, which is implemented as a three-layer CNN, to extract a sequence-based protein representation. In the second stage, each drug compound in the behavioral sequence of the target protein is independently encoded through the same GCN-based drug encoder, yielding a set of behavior-aware drug representations that collectively reflect the interaction context of the target protein. In the third stage, these behavioral representations are weighted by their associated interaction strengths and passed through a multi-head attention mechanism, which adaptively aggregates the behavioral context by attending to the most informative interaction partners. In the fourth stage, the refined behavioral representation, the drug representation, and the protein representation are projected into an embedding space through alignment layers and concatenated to form a unified interaction feature vector, which is subsequently fed into a classification network to produce the final binary interaction prediction.

The key design principle of this framework is that the behavioral sequence of a target protein encodes implicit pharmacological preferences that cannot be inferred from the protein sequence alone. By explicitly modeling this behavioral context and integrating it with pairwise drug–protein features, the proposed framework enables more informed and context-aware interaction prediction beyond the conventional pairwise matching paradigm. In addition, unlike BiBLDR [39], we do not explicitly introduce predefined similarity scores during behavioral sequence construction. Instead, the query drug and behavioral drugs are encoded by the same GCN encoder, which preserves richer molecular structural

features than a single handcrafted similarity value. These structural representations carry more semantic information for DTI prediction, while the attention-based fusion module further learns the relevance of historical behavioral drugs in a task-adaptive manner. This design avoids relying on fixed, manually defined similarity functions and is more suitable for our protein-centric DTI prediction setting.

3.2. Drug and target representation

Each drug is represented as a molecular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where nodes denote atoms, edges denote chemical bonds, and each atom v_i has a physicochemical feature vector $\mathbf{x}_i \in \mathbb{R}^{d_a}$. For mini-batch processing, graphs are padded to $N_{\max}$ nodes using distinguishable virtual nodes, and self-loops are added to support information propagation. A GNN iteratively aggregates neighboring features through message passing, enabling each atom representation to capture both its local bonding environment and increasingly long-range molecular dependencies. To implement this process, we adopt a GCN as the molecular encoder and first project the initial atomic features into a d -dimensional embedding space, as shown in Eq (3.1):

$$\mathbf{h}_i^{(0)} = \mathbf{W}_{init}\mathbf{x}_i, \quad (3.1)$$

where $\mathbf{x}_i \in \mathbb{R}^{d_a}$ is the initial node feature and $\mathbf{W}_{init} \in \mathbb{R}^{d \times d_a}$ is a learnable projection matrix. The virtual-node dimension is fixed at zero to prevent padding from affecting the learned representations. The embeddings are then refined through L degree-normalized GCN layers, producing $\mathbf{H}^{(L)} \in \mathbb{R}^{N_{\max} \times d}$, which captures local and long-range molecular structural information and is retained without global pooling for subsequent interaction modeling.

Each target protein sequence is truncated or padded to a fixed length L_p and mapped to d -dimensional residue embeddings, with index zero reserved for padding. A three-layer 1D-CNN, composed of convolution, rectified linear unit (ReLU), and batch-normalization blocks with progressively smaller kernels and strided compression, is then used to capture local sequence motifs at multiple scales. The resulting position-wise protein representation has shape $L'_p \times d$, where L'_p is the compressed sequence length, and is retained without global pooling for subsequent feature fusion.

3.3. Protein behavioral sequence modeling

The central hypothesis of this work is that the binding preferences of a target protein can be characterized not only by its amino acid sequence but also by the collection of drug compounds with which it has been observed to interact in known interaction data. Motivated by this intuition, we define the behavioral sequence of a target protein as the set of drug compounds in the training set that share the same target protein, together with their corresponding experimentally measured interaction labels. Formally, let $\mathcal{D}_{train}$ denote the training set consisting of triplets (d_i, p_i, y_i) , where d_i is a drug compound, p_i is a target protein, and $y_i \in \{0, 1\}$ is the binary interaction label. For a query sample (d_q, p_q, y_q) , the sample-specific behavioral record of protein p_q is defined in Eq (3.2).

$$\mathcal{B}(p_q, d_q) = \{(d_i, y_i) \mid (d_i, p_i, y_i) \in \mathcal{D}_{train}, p_i = p_q, d_i \neq d_q\}. \quad (3.2)$$

To prevent label leakage, the query interaction (d_q, p_q, y_q) is excluded from the behavioral record of p_q . Since the number of historical interactions varies across proteins, the remaining record $\mathcal{B}(p_q, d_q)$ is divided into non-overlapping chunks of fixed length K , each forming a behavioral sequence instance.

This strategy enables fixed-size modeling and exposes the model to diverse subsets of interaction history without revealing the query label.

3.3.1. Behavioral drug encoding

Given a behavioral sequence chunk $\mathcal{B}_k(p_q, d_q) = \{(d_1, y_1), (d_2, y_2), \dots, (d_K, y_K)\}$ associated with target protein p_q , each behavioral drug d_j is independently encoded into a node-level feature matrix using the same GCN module as that used for the query drug. By encoding both the query drug and the behavioral drugs with a unified molecular encoder, the resulting representations lie in the same embedding space, ensuring that structural similarities between the query compound and historical binding partners are geometrically meaningful. For each behavioral drug d_j , the output of the GCN is a node-level feature matrix $\mathbf{H}_j \in \mathbb{R}^{N_{\max} \times d}$, which is then flattened into a vector $\tilde{\mathbf{h}}_j$. The compact behavioral embedding of the j -th behavioral drug, denoted by $\mathbf{b}_j$, is obtained by projecting its flattened molecular representation $\tilde{\mathbf{h}}_j$ through a shared ReduceLayer.

The interaction labels y_j associated with each behavioral drug carry important pharmacological information, with a positive label indicating confirmed binding and a negative label indicating non-interaction. To enable the model to learn effective behavioral supervision signals from historical interactions, we incorporate the interaction labels as explicit weighting signals to modulate the contribution of each behavioral embedding. Specifically, the interaction label y_j is passed through an exponential function to produce a non-negative scalar weight, as shown in Eq (3.3).

$$\hat{y}_j = \exp(y_j). \quad (3.3)$$

The exponential transformation is preferred over direct binary multiplication because it does not completely eliminate the representations associated with non-interacting drugs. Thus, confirmed interactions are emphasized, while non-interacting drugs are retained as useful contrastive context for characterizing the behavioral profile of the target. The positive and smooth weighting scheme also avoids abrupt information loss before feature aggregation. The behavioral embedding $\mathbf{b}_j$ is then scaled by its corresponding weight, as shown in Eq (3.4):

$$\hat{\mathbf{b}}_j = \hat{y}_j \cdot \mathbf{b}_j. \quad (3.4)$$

This weighting mechanism assigns higher importance to behavioral drugs with confirmed interactions while preserving the complementary information provided by non-interacting compounds. The behavioral context matrix $\hat{\mathbf{B}} = [\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \dots, \hat{\mathbf{b}}_K]^T$ is subsequently passed to the attention-based fusion module described in the following section.

3.4. Behavior-aware feature fusion

Given the weighted behavioral context matrix $\hat{\mathbf{B}} \in \mathbb{R}^{K \times d_b}$ produced by the preceding module, the goal of this stage is to refine the behavioral representations by modeling the interdependencies among the K behavioral drugs and selectively aggregating the most informative contextual signals. To this end, we design a behavior-aware feature fusion module that employs a multi-head self-attention mechanism to capture dependencies among behavioral representations.

The multi-head attention mechanism allows the model to jointly attend to information from different representation subspaces at different positions and has demonstrated strong performance in capturing

relational dependencies in sequential data. In our framework, the K weighted behavioral embeddings are treated as a sequence, and self-attention is applied so that each behavioral drug can attend to all other behavioral drugs within the same behavioral context. Formally, the multi-head attention output is computed as in Eq (3.5), where each attention head is defined in Eq (3.6) and the attention operation is given in Eq (3.7).

$$\mathbf{MHA}(\hat{\mathbf{B}}) = \text{Concat}(\mathbf{head}_1, \dots, \mathbf{head}_H) \mathbf{W}^O, \quad (3.5)$$

where each attention head is computed as:

$$\mathbf{head}_h = \text{Attention}(\hat{\mathbf{B}} \mathbf{W}_h^Q, \hat{\mathbf{B}} \mathbf{W}_h^K, \hat{\mathbf{B}} \mathbf{W}_h^V), \quad (3.6)$$

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q} \mathbf{K}^\top}{\sqrt{d_b/H}}\right) \mathbf{V}. \quad (3.7)$$

where $\hat{\mathbf{B}} \in \mathbb{R}^{K \times d_b}$ denotes the weighted behavioral context matrix and H is the number of attention heads. Here, $\mathbf{W}^O$ denotes the output projection matrix, while $\mathbf{W}_h^Q$, $\mathbf{W}_h^K$, and $\mathbf{W}_h^V$ denote the learnable query, key, and value projection matrices for the h -th head, respectively. Accordingly, $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ denote the query, key, and value matrices obtained from these linear projections, $\mathbf{head}_h$ denotes the output of the h -th attention head, and $\mathbf{MHA}(\hat{\mathbf{B}})$ denotes the concatenated multi-head attention output.

To preserve the behavioral signals produced by the preceding stage, a residual connection is introduced so that the weighting information is not lost during attention-based aggregation, and the resulting representation is normalized by a batch normalization layer to obtain $\mathbf{M}$. The normalized attention output is then fed into a position-wise feed-forward layer for further transformation, yielding an intermediate representation $\mathbf{Z}$ through a linear mapping, as shown in Eq (3.8). After that, a second residual connection combines the feed-forward output with the multi-head attention branch, and the combined representation is further normalized by another batch normalization layer, as shown in Eq (3.9).

$$\mathbf{Z} = \mathbf{W}_f \cdot \mathbf{M} + \mathbf{c}_f, \quad (3.8)$$

$$\mathbf{R} = \text{BN}_2(\mathbf{MHA}(\hat{\mathbf{B}}) + \text{Dropout}_2(\mathbf{Z})). \quad (3.9)$$

To generate the final interaction prediction, the drug molecular and protein sequence representations are independently mapped to a common d_{align} -dimensional space through two-layer alignment networks with batch normalization, ReLU activation, and dropout, producing $\mathbf{a}_d, \mathbf{a}_p \in \mathbb{R}^{d_{\text{align}}}$. The attention-refined behavioral representation $\mathbf{R} \in \mathbb{R}^{K \times d_b}$ is flattened as $\mathbf{a}_b \in \mathbb{R}^{K \cdot d_b}$, and the three representations are concatenated into a unified feature vector. This vector is passed through a multilayer classifier with progressively decreasing hidden dimensions, batch normalization, ReLU activation, and dropout, followed by a softmax layer that outputs the binary interaction probability.

4. Results

4.1. Datasets and evaluation metrics

We evaluated the performance of the proposed method on three widely used public benchmark datasets for DTI prediction: BindingDB [40], BioSNAP [30], and Human [41]. These datasets consist

of small-molecule drugs and proteins, and each drug–protein pair is annotated with a binary interaction label (0 or 1). BindingDB is an experimentally validated database of drug–protein binding affinities, and we adopted the low-bias preprocessed version described in [32]. BioSNAP is derived from DrugBank and contains 4510 drugs and 2181 proteins, in which the confirmed positive pairs are balanced by an equal number of randomly sampled negative pairs. Human contains highly credible negative samples constructed through a physicochemical dissimilarity-based screening protocol, and we used its balanced version with equal numbers of positive and negative interactions. The statistics of the three datasets are summarized in Table 1.

Table 1. Statistics of the three benchmark datasets.

Dataset	Drugs	Proteins	Interactions
BindingDB	14,643	2623	49,199
BioSNAP	4510	2181	27,464
Human	2726	2001	6728

To comprehensively evaluate the predictive performance of the proposed model on the binary DTI classification task, we adopted four widely used evaluation metrics: the area under the receiver operating characteristic curve (AUROC), the area under the precision–recall curve (AUPRC), accuracy (ACC), and the F1-score. ACC measures the overall proportion of correctly classified samples among all predictions, whereas the F1-score provides a balanced assessment of precision and recall, making it particularly informative when the class distribution is imbalanced. AUROC evaluates the model’s ability to distinguish positive and negative samples across different classification thresholds, while AUPRC focuses on the precision–recall trade-off and is especially informative for assessing positive-interaction prediction performance. Let true positives, true negatives, false positives, and false negatives be denoted by TP , TN , FP , and FN , respectively. The ACC is defined in Eq (4.1):

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (4.1)$$

where a higher ACC indicates that the model correctly predicts a larger proportion of drug–target pairs. However, ACC alone may be insufficient when the class distribution is imbalanced.

Precision reflects the reliability of predicted positive interactions, that is, the proportion of drug–target pairs predicted to interact that are truly positive. Recall reflects the coverage of actual positive interactions, that is, the proportion of truly interacting drug–target pairs that are correctly identified by the model. Based on these two complementary perspectives, the F1-score is defined as the harmonic mean of precision and recall. Precision, recall, and the F1-score are given in Eqs (4.2)–(4.4):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4.2)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (4.3)$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (4.4)$$

A higher F1-score indicates a better balance between precision and recall, reflecting stronger overall performance of the model in identifying drug-target interactions. Higher AUROC and AUPRC values further indicate stronger threshold-independent discrimination ability and more reliable positive-interaction ranking performance, respectively. Therefore, the combination of AUROC, AUPRC, ACC, and the F1-score enables a more comprehensive and reliable evaluation of model performance.

4.2. Implementation details

All experiments were implemented in Python using the PyTorch deep learning framework and were conducted on a single NVIDIA RTX 4070 GPU. The maximum number of nodes per molecular graph was set to $N_{\max} = 290$, and protein sequences were truncated or zero-padded to a fixed length of $L_p = 1200$. The GCN encoder consisted of $L = 3$ graph convolutional layers with a hidden embedding dimension of $d = 128$. The number of attention heads in the multi-head self-attention module was set to $H = 8$, the behavioral embedding dimension was set to $d_b = 512$, and the behavioral sequence chunk size was set to $K = 4$. The intermediate hidden dimension of the ReduceLayer was set to $d_{\text{mid}} = 4096$, and the alignment layers projected the drug and protein representations into a shared space of dimension $d_{\text{align}} = 1024$. A dropout rate of 0.2 was applied consistently across all modules.

The model was optimized using the Adam optimizer with an initial learning rate of 1×10^{-3} . A step-based learning rate decay schedule was adopted, reducing the learning rate by a factor of 0.1 every 10 epochs. All models were trained for 100 epochs with a mini-batch size of 32. For each dataset, we randomly split the samples into a training set and a test set at a ratio of 8:2. All reported results represent the average performance over five independent runs with different random seeds to ensure statistical reliability. To further assess whether the observed improvements are statistically significant, we conducted two-sided Welch's t -tests based on the five independent runs. For each dataset and evaluation metric, the proposed method was compared with the strongest-performing baseline, and $p < 0.05$ was considered statistically significant.

4.3. Comparison with baseline methods

To evaluate the effectiveness of the proposed framework, we compared it against nine representative baseline methods on three benchmark datasets under a random-split setting. These methods were selected to cover distinct paradigms in DTI prediction. Specifically, support vector machine (SVM) [42] and random forest (RF) [43] were included as classical machine-learning baselines based on handcrafted features. Graph drug-target affinity (GraphDTA) [44] employs GNNs to encode drug molecular structures together with sequence-based protein representations. Transformer compound-protein interaction (TFCPI) [45] introduces transformer-based self-attention to model compound-protein interactions, while MolTrans [30] captures interactions between drug and protein substructures through a Transformer architecture. DrugBAN [32] uses bilinear attention to explicitly model local drug-target interactions, CAT-DTI [46] further exploits cross-attention and Transformer modules for feature interaction, and hyper graph convolutional network for DTI (HyperGCN-DTI) [47] uses graph convolutional networks to integrate multimodal drug and protein features for DTI prediction. In addition, graph transformer for DTI (GraphormerDTI) [48] was included as a recent graph-transformer-based method that captures both molecular topology and long-range structural dependencies. Together, these baselines provide a comprehensive comparison across classical machine learning, GNNs, attention-based models, and

modern transformer architectures. All results are reported in terms of AUROC, AUPRC, ACC, and the F1-score and are averaged over five independent runs. The complete results are presented in Table 2.

Table 2. Comparison with baseline methods on three benchmark datasets in terms of AUROC, AUPRC, ACC, and F1-score. The number in parentheses denotes the number of trainable parameters. * and ** indicate that the improvement of our method over the strongest baseline for the corresponding metric is statistically significant at $p < 0.05$ and $p < 0.01$, respectively, according to a two-sided Welch's t -test.

Dataset	Method	AUROC	AUPRC	ACC	F1
BindingDB	SVM	0.904 $\pm$ 0.009	0.865 $\pm$ 0.012	0.824 $\pm$ 0.015	0.785 $\pm$ 0.018
	RF	0.942 $\pm$ 0.011	0.923 $\pm$ 0.016	0.871 $\pm$ 0.018	0.844 $\pm$ 0.021
	GraphDTA (4.75M)	0.944 $\pm$ 0.009	0.923 $\pm$ 0.012	0.874 $\pm$ 0.017	0.880 $\pm$ 0.019
	TFCPI (0.47M)	0.947 $\pm$ 0.008	0.932 $\pm$ 0.011	0.888 $\pm$ 0.014	0.888 $\pm$ 0.016
	MolTrans (62.86M)	0.947 $\pm$ 0.010	0.927 $\pm$ 0.013	0.884 $\pm$ 0.018	0.886 $\pm$ 0.017
	DrugBAN (1.07M)	0.961 $\pm$ 0.008	0.948 $\pm$ 0.010	0.901 $\pm$ 0.012	0.903 $\pm$ 0.014
	CAT-DTI (2.37M)	0.960 $\pm$ 0.009	0.947 $\pm$ 0.011	0.896 $\pm$ 0.016	0.900 $\pm$ 0.013
	HyperGCN-DTI (0.85M)	0.958 $\pm$ 0.007	0.950 $\pm$ 0.012	0.892 $\pm$ 0.015	0.897 $\pm$ 0.016
	GraphormerDTI (6.92M)	0.962 $\pm$ 0.008	0.949 $\pm$ 0.010	0.900 $\pm$ 0.013	0.902 $\pm$ 0.015
	Ours (8.32M)	0.983* $\pm$ 0.006	0.966* $\pm$ 0.008	0.947** $\pm$ 0.009	0.957** $\pm$ 0.007
BioSNAP	SVM	0.819 $\pm$ 0.021	0.839 $\pm$ 0.019	0.750 $\pm$ 0.026	0.827 $\pm$ 0.024
	RF	0.857 $\pm$ 0.018	0.872 $\pm$ 0.021	0.793 $\pm$ 0.024	0.787 $\pm$ 0.028
	GraphDTA (4.75M)	0.871 $\pm$ 0.019	0.870 $\pm$ 0.022	0.800 $\pm$ 0.025	0.807 $\pm$ 0.023
	TFCPI (0.47M)	0.876 $\pm$ 0.020	0.881 $\pm$ 0.018	0.797 $\pm$ 0.029	0.803 $\pm$ 0.024
	MolTrans (62.86M)	0.895 $\pm$ 0.013	0.899 $\pm$ 0.016	0.820 $\pm$ 0.021	0.825 $\pm$ 0.019
	DrugBAN (1.07M)	0.902 $\pm$ 0.011	0.905 $\pm$ 0.014	0.836 $\pm$ 0.018	0.838 $\pm$ 0.017
	CAT-DTI (2.37M)	0.909 $\pm$ 0.010	0.907 $\pm$ 0.013	0.836 $\pm$ 0.019	0.840 $\pm$ 0.016
	HyperGCN-DTI (0.85M)	0.911 $\pm$ 0.013	0.914 $\pm$ 0.010	0.842 $\pm$ 0.015	0.848 $\pm$ 0.012
	GraphormerDTI (6.92M)	0.906 $\pm$ 0.012	0.908 $\pm$ 0.011	0.838 $\pm$ 0.017	0.841 $\pm$ 0.015
	Ours (8.32M)	0.932* $\pm$ 0.008	0.929* $\pm$ 0.009	0.868* $\pm$ 0.011	0.923** $\pm$ 0.010
Human	SVM	0.913 $\pm$ 0.015	0.905 $\pm$ 0.018	0.838 $\pm$ 0.022	0.811 $\pm$ 0.025
	RF	0.939 $\pm$ 0.012	0.927 $\pm$ 0.014	0.866 $\pm$ 0.019	0.848 $\pm$ 0.021
	GraphDTA (4.75M)	0.965 $\pm$ 0.009	0.955 $\pm$ 0.011	0.908 $\pm$ 0.014	0.907 $\pm$ 0.016
	TFCPI (0.47M)	0.954 $\pm$ 0.011	0.941 $\pm$ 0.013	0.879 $\pm$ 0.018	0.891 $\pm$ 0.017
	MolTrans (62.86M)	0.981 $\pm$ 0.007	0.976 $\pm$ 0.009	0.941 $\pm$ 0.012	0.943 $\pm$ 0.011
	DrugBAN (1.07M)	0.981 $\pm$ 0.008	0.969 $\pm$ 0.010	0.940 $\pm$ 0.013	0.940 $\pm$ 0.012
	CAT-DTI (2.37M)	0.983 $\pm$ 0.006	0.976 $\pm$ 0.008	0.942 $\pm$ 0.011	0.944 $\pm$ 0.010
	HyperGCN-DTI (0.85M)	0.980 $\pm$ 0.010	0.977 $\pm$ 0.011	0.938 $\pm$ 0.015	0.941 $\pm$ 0.012
	GraphormerDTI (6.92M)	0.982 $\pm$ 0.007	0.974 $\pm$ 0.009	0.941 $\pm$ 0.012	0.943 $\pm$ 0.011
	Ours (8.32M)	0.992* $\pm$ 0.004	0.986 $\pm$ 0.006	0.975** $\pm$ 0.007	0.986** $\pm$ 0.005

The results in Table 2 demonstrate that the proposed method achieves the strongest overall performance across all three benchmark datasets in terms of AUROC, AUPRC, ACC, and the F1-score. On BindingDB, our method obtains an AUROC of 0.983, an AUPRC of 0.966, an ACC of 0.947, and an F1-score of 0.957. GraphormerDTI provides the strongest baseline AUROC, HyperGCN-DTI achieves the best baseline AUPRC, and DrugBAN achieves the best baseline ACC and F1-score. Compared with the best baseline result for each metric, the proposed model improves AUROC, AUPRC, ACC, and the F1-score by 0.021, 0.016, 0.046, and 0.054, respectively. Compared specifically with GraphormerDTI, our method yields improvements of 0.021, 0.017, 0.047, and 0.055 in the four metrics, respectively. These improvements indicate that the proposed framework not only enhances classification accuracy but also provides better threshold-independent discrimination and more reliable positive-interaction ranking performance on large-scale DTI data.

On BioSNAP, our method also consistently outperforms all baselines, achieving an AUROC of 0.932, an AUPRC of 0.929, an ACC of 0.868, and an F1-score of 0.923. HyperGCN-DTI is the strongest baseline across all four metrics. Compared with its results, the proposed model improves AUROC, AUPRC, ACC, and the F1-score by 0.021, 0.015, 0.026, and 0.075, respectively. Relative to GraphormerDTI, the corresponding improvements are 0.026, 0.021, 0.030, and 0.082. The particularly clear gain in the F1-score, together with the improvements in AUROC and AUPRC, suggests that the model is more effective at identifying positive interactions while maintaining strong overall discrimination ability across decision thresholds. On the Human dataset, our method achieves the best performance, with an AUROC of 0.992, an AUPRC of 0.986, an ACC of 0.975, and an F1-score of 0.986. CAT-DTI provides the strongest baseline AUROC, ACC, and F1-score, whereas HyperGCN-DTI achieves the best baseline AUPRC. Compared with the best baseline result for each metric, these values represent improvements of 0.009 in AUROC, 0.009 in AUPRC, 0.033 in ACC, and 0.042 in the F1-score. Moreover, our method outperforms GraphormerDTI by 0.010, 0.012, 0.034, and 0.043 in AUROC, AUPRC, ACC, and the F1-score, respectively. Since the Human dataset contains highly credible negative samples and is generally considered more challenging, the consistent improvements across all four metrics further verify the robustness and generalization capability of the proposed framework.

In addition to predictive performance, Table 2 also reports the trainable parameter counts of the deep learning baselines. Our model contains 8.32M parameters, which is larger than lightweight models such as TFCPI (0.47M), HyperGCN-DTI (0.85M), DrugBAN (1.07M), and CAT-DTI (2.37M), but substantially smaller than MolTrans (62.86M). GraphormerDTI contains 6.92M parameters, making its model size the closest to that of our method among the compared baselines. Despite using only 1.40M additional parameters, our method consistently outperforms GraphormerDTI across all datasets and metrics, indicating that the gains cannot be attributed solely to increased model size. Compared with MolTrans, our method also achieves consistently better results while using only a small fraction of its parameters, suggesting a more favorable balance between model capacity and predictive effectiveness. The additional parameters in our framework mainly support behavior-aware contextual modeling, which provides richer interaction features rather than merely increasing model complexity.

Overall, the experimental results show that our method outperforms both traditional machine-learning baselines and recent deep learning approaches across all datasets and evaluation metrics while maintaining a moderate parameter scale. Classical machine-learning methods such as SVM and RF achieve relatively limited performance, indicating that shallow handcrafted representations are insufficient for modeling complex drug-target relationships. Deep learning baselines, including GraphDTA, TFCPI, MolTrans, DrugBAN, CAT-DTI, HyperGCN-DTI, and GraphormerDTI, achieve stronger results by learning task-oriented representations from molecular graphs and protein sequences, but they remain inferior to our method. In particular, the consistent improvements over HyperGCN-DTI and GraphormerDTI demonstrate that multimodal feature integration or the modeling of molecular topology and long-range dependencies alone is insufficient to fully capture the contextual interaction patterns of target proteins. We attribute the advantage of our method to the effective integration of drug molecular information, protein sequence semantics, and behavior-aware contextual representations, which enables the model to learn more comprehensive and informative interaction features. Therefore, the proposed framework demonstrates strong effectiveness, robustness, and generalization capability in DTI prediction.

Statistical testing further supports the above comparisons. On BindingDB, the improvements

over the strongest baseline are significant for AUROC ($p = 0.0019$), AUPRC ($p = 0.0423$), ACC ($p < 0.001$), and F1-score ($p < 0.001$). On BioSNAP, significant differences are also observed for AUROC ($p = 0.0191$), AUPRC ($p = 0.0377$), ACC ($p = 0.0157$), and F1-score ($p < 0.001$). On the Human dataset, the improvements are significant for AUROC ($p = 0.0270$), ACC ($p < 0.001$), and F1-score ($p < 0.001$). These results indicate that most of the observed gains over the strongest competing methods are statistically supported, while also avoiding an overstatement of the AUPRC improvement on the Human dataset.

4.4. Ablation study

To further investigate the contributions of the key modules in the proposed framework, we conducted ablation experiments on the BindingDB, BioSNAP, and Human datasets. Specifically, we evaluated four simplified variants: one without the behavioral sequence, one without the attention mechanism, one without exponential label transformation, and one in which multi-head attention was replaced with mean pooling. The corresponding results are presented in Table 3.

Table 3. Ablation study on three benchmark datasets. BS denotes behavioral sequence, AM denotes attention mechanism, ELT denotes exponential label transformation, and MX denotes replacing multi-head attention with mean pooling.

Variant	BindingDB				BioSNAP				Human			
	AUROC	AUPRC	ACC	F1	AUROC	AUPRC	ACC	F1	AUROC	AUPRC	ACC	F1
w/o BS (0.95M)	0.957	0.943	0.904	0.912	0.886	0.891	0.821	0.829	0.970	0.963	0.928	0.941
w/o AM	0.973	0.958	0.912	0.938	0.898	0.903	0.833	0.835	0.982	0.978	0.940	0.967
w/o ELT	0.980	0.962	0.943	0.953	0.928	0.925	0.863	0.918	0.989	0.983	0.971	0.982
MX	0.971	0.955	0.910	0.934	0.895	0.899	0.829	0.831	0.979	0.975	0.936	0.963
Full model (8.32M)	0.983	0.966	0.947	0.957	0.932	0.929	0.868	0.923	0.992	0.986	0.975	0.986

The ablation results in Table 3 show consistent performance trends across the three datasets. Despite substantially reducing the parameter count by eliminating the behavioral modeling branch, removing the behavioral sequence causes the largest overall performance degradation, confirming that this branch provides essential contextual information beyond the intrinsic features of drugs and proteins rather than merely increasing model capacity. Removing the attention mechanism also reduces all evaluation metrics, indicating that selectively modeling dependencies among behavioral representations is important for effective feature aggregation. Notably, MX performs even worse than the variant without the attention mechanism. This suggests that average pooling compresses all behavioral representations into a uniform summary, which may discard interaction-specific features and dependencies among behavioral drugs. In contrast, multi-head attention preserves more informative contextual signals by adaptively modeling and weighting different behavioral representations.

The variant without exponential label transformation remains close to the full model but performs consistently worse. This result suggests that directly using binary labels preserves useful interaction information, while exponential transformation provides a more expressive weighting of behavioral representations. Overall, the full model achieves the best results by jointly exploiting behavioral context, label-aware weighting, and attention-based aggregation, demonstrating that these components play complementary roles in DTI prediction.

4.5. Cold-start evaluation

To further assess the generalization ability of the proposed framework, we conducted an unseen-target cold-start evaluation on the BindingDB dataset, in which the target proteins in the test set do not appear in the training set. The results are summarized in Table 4.

Table 4. Performance comparison under the unseen-target cold-start setting on the BindingDB dataset.

Method	AUROC	AUPRC	ACC	F1
GraphDTA (4.75M)	0.752 ± 0.021	0.821 ± 0.023	0.701 ± 0.023	0.798 ± 0.021
TFCPI (0.47M)	0.662 ± 0.030	0.749 ± 0.033	0.628 ± 0.031	0.651 ± 0.039
MolTrans (62.86M)	0.816 ± 0.018	0.853 ± 0.020	0.747 ± 0.020	0.836 ± 0.019
DrugBAN (1.07M)	0.875 ± 0.016	0.918 ± 0.017	0.794 ± 0.018	0.821 ± 0.020
CAT-DTI (2.37M)	0.851 ± 0.018	0.889 ± 0.020	0.773 ± 0.020	0.807 ± 0.019
GraphormerDTI (6.92M)	0.829 ± 0.019	0.870 ± 0.021	0.756 ± 0.021	0.796 ± 0.022
Ours (8.32M)	<u>0.861 ± 0.017</u>	<u>0.901 ± 0.019</u>	0.798 ± 0.019	0.816 ± 0.018

As shown in Table 4, the performance of our method decreases under the unseen-target cold-start setting compared with the conventional random-split setting. The main reason is that the test targets do not appear in the training set and therefore lack historical interaction records, making their behavioral sequences unavailable and preventing the behavior-aware module from fully exploiting interaction-context information. Nevertheless, our method ranks second in both AUROC and AUPRC, achieves the best ACC, and maintains a competitive F1-score. Compared with GraphormerDTI, our method improves AUROC, AUPRC, ACC, and the F1-score by 0.032, 0.031, 0.042, and 0.020, respectively. These results demonstrate that, even without the behavioral sequence, the complementary drug molecular structure and protein sequence representations still enable the proposed framework to retain strong predictive performance and exhibit good generalization ability for previously unseen targets. They also suggest that constructing informative behavioral priors for targets without interaction histories is an important direction for future improvement.

4.6. Hyperparameter sensitivity analysis

To examine the robustness of the proposed framework, we analyzed the influence of two key hyperparameters, namely the behavioral sequence chunk size $K \in [2, 4, 6, 8]$ and the behavioral embedding dimension $d_b \in [64, 128, 256, 512]$. Their effects on AUROC and AUPRC are illustrated in Figure 2.

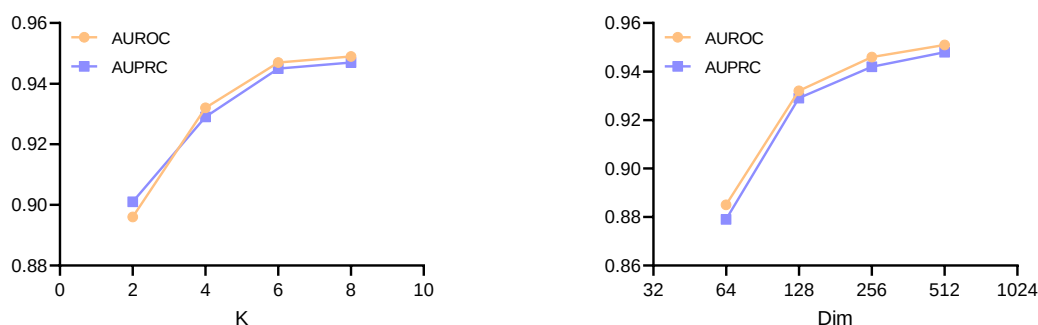


Figure 2. Sensitivity of AUROC and AUPRC to the behavioral sequence chunk size K (left) and behavioral embedding dimension (right).

As shown in Figure 2, both AUROC and AUPRC improve as K increases, with a substantial gain from 2 to 4 and progressively smaller improvements beyond 6. This indicates that a longer behavioral sequence provides richer interaction context, although its benefit gradually saturates as additional historical drugs contribute less complementary information. A similar trend is observed for d_b : Performance increases markedly from 64 to 128 and then improves more gradually up to 512. These results show that sufficient embedding capacity is important for preserving behavioral information, while excessively large settings offer limited additional gains and increase computational cost.

4.7. Real-world scenario validation

To evaluate the applicability of the proposed model in real-world scenarios, we further analyzed its predictions for unannotated drug-target pairs, with particular attention to pairs assigned high prediction scores. In practical drug discovery, many potential drug-target combinations have not yet been experimentally characterized. Therefore, these unannotated samples may include true but previously undiscovered biological interactions, and the high-scoring pairs identified by our model may provide useful candidates for subsequent experimental validation.

Specifically, we selected the 10 unannotated drug-target pairs with the highest prediction scores and queried them against PubChem BioAssay. We then examined whether existing experimental bioactivity records or curated evidence supported the predicted associations. In PubChem BioAssay, records labeled as *Active* or *Probe* were treated as supporting evidence for the predicted interactions. The *Active* label indicates a clear experimentally observed response, whereas *Probe* denotes a compound with experimentally confirmed activity against the target that has been curated as a useful chemical probe for investigating its biological function. The validation results for the 10 highest-ranked unannotated drug-target pairs are summarized in Table 5.

Table 5. PubChem BioAssay validation results for the 10 highest-ranked unannotated drug-target pairs. Gray-shaded rows indicate predicted pairs supported by PubChem BioAssay records.

Rank	Drug	Target	BioAssay ID	Type	Evidence
1	3076090	P21396	125713	Probe	[49]
2	122182317	Q02127	1234866	Probe	[50]
3	10383508	Q12884	265517	Probe	[51]
4	78322705	P07339	–	–	–
5	135400189	O00408	735486	Active	[52]
6	118979215	O60674	1798284	Active	[53]
7	86288088	Q01469	–	–	–
8	5281614	P0DUB6	404692	Probe	[54]
9	176155	P00533	624997	Active	[55]
10	90415325	P07339	–	–	–

As shown in Table 5, seven of the 10 highest-ranked unannotated drug-target pairs are supported by PubChem BioAssay records, including both *Active* and *Probe* annotations. This result indicates that the

high-confidence predictions generated by our model are not merely numerical artifacts but are consistent with external experimental or curated bioactivity evidence. In particular, the presence of multiple supported pairs among the highest-ranked candidates suggests that the proposed framework can prioritize biologically meaningful drug-target associations from the unannotated candidate space. Although several predicted pairs remain unsupported by current PubChem records, they may still represent plausible candidates for future experimental investigation because the absence of database evidence does not necessarily imply the absence of a true interaction. Overall, this real-world validation further demonstrates the practical potential of the proposed model for discovering previously unannotated DTIs.

5. Discussion

Several methodological and practical issues of the proposed framework warrant further discussion. First, the use of historical interaction labels should be distinguished from information leakage. In behavior-sequence learning, historical interactions are commonly used as contextual information to characterize entity preferences and predict unseen interactions [56–58]. Our framework follows this paradigm: The query drug–target pair is explicitly excluded from the behavioral sequence, and the behavioral context is constructed exclusively from interactions available in the training set. Therefore, neither the query label nor any test-set labels are exposed to the model. The retained historical interaction labels instead represent prior behavioral information of the target protein, analogous to historical interactions used in sequential recommendation and behavior-based association-prediction methods such as BiBLDR [39]. Nevertheless, such prior information may provide stronger predictive signals for targets with richer interaction histories. Accordingly, the reported performance should primarily be interpreted under a context-aware warm-start setting in which historical interactions of the target protein are available.

Second, our framework adopts a lightweight three-layer CNN rather than a pretrained protein language model for protein representation. Although pretrained models such as ESM and ProtT5 can provide more expressive protein representations, they generally introduce substantially larger parameter counts and higher computational and memory costs. The CNN encoder keeps the overall framework computationally tractable and allows the contribution of the proposed behavior-aware modeling strategy to be evaluated without relying heavily on large-scale pretrained representations. This design therefore reflects a trade-off between representation capacity and computational efficiency.

Finally, the practical value of computational DTI prediction depends not only on benchmark performance but also on whether high-confidence predictions are supported by independent biological evidence. Retrospective validation using public bioactivity databases is commonly employed in computational drug discovery to assess the biological plausibility of predicted associations and prioritize candidates for subsequent investigation [59, 60]. PubChem BioAssay, in particular, contains experimentally derived bioactivity records from large-scale screening and medicinal chemistry studies and therefore provides an independent source of supporting evidence [61]. The agreement between our top-ranked predictions and PubChem BioAssay records provides additional support for the reliability of the proposed framework as a candidate-prioritization tool. Nevertheless, such retrospective database confirmation should be interpreted as external supporting evidence rather than prospective prediction or direct experimental validation.

6. Conclusions

In this work, we propose a novel framework for drug-target interaction prediction that integrates drug molecular structure, protein sequence information, and behavior-aware contextual representations within a unified learning architecture. By combining graph-based drug encoding, sequence-based protein representation learning, interaction-guided behavioral modeling, and attention-based feature fusion, the proposed framework captures complementary information from multiple perspectives and generates more discriminative representations for DTI prediction. Experiments on the BindingDB, BioSNAP, and Human datasets demonstrate that the proposed framework consistently outperforms representative baseline methods, while the ablation study confirms the effectiveness of the behavioral sequence and attention modules.

Overall, the proposed framework provides an effective computational solution for DTI prediction and offers promising support for accelerating early-stage drug discovery. Future DTI research is expected to move toward multimodal foundation models that integrate molecular structures, protein sequences, three-dimensional conformations, omics data, and biomedical knowledge, as well as trustworthy prediction systems with calibrated uncertainty and mechanistic explanations. We will continue to refine and extend our framework toward these goals. Another practical challenge is the severe imbalance between known and unknown DTIs. Validated interactions represent only a small fraction of possible drug-target pairs, while most pairs remain unannotated and cannot be treated as true negatives. Because balanced benchmarks do not fully reflect this sparse real-world distribution, future work will explore positive-unlabeled learning, improved negative sampling, and training methods for imbalanced data in large-scale candidate screening. In addition, inspired by recent computational studies on small-molecule-miRNA associations [62–64], we plan to extend the framework to predict small-molecule-miRNA interactions by incorporating miRNA-specific sequence and structural representations and their interaction history.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors sincerely thank all reviewers for their valuable comments and constructive suggestions. This work was partially supported by the Innovation Research Foundation of National University of Defense Technology (Innovation Research Foundation of NUDT) under Grant No. ZK23-51, the Fundamental Research Funds for the Central Universities under Grant No. B250201270, the Outstanding Youth Project of the Hunan Provincial Department of Education under Grant No. 24B0865, the National Natural Science Foundation of China under Grant No. 62306323, the Science and Technology Program of Hunan Province under Grant No. 2024RC3117, and the Outstanding Youth Project of the Hunan Provincial Department of Education under Grant No. 25B0858.

References

1. D. Sun, W. Gao, H. Hu, S. Zhou, Why 90% of clinical drug development fails and how to improve it?, *Acta Pharm. Sin. B*, **12** (2022), 3049–3062. <https://doi.org/10.1016/j.apsb.2022.02.002>

2. A. Schuhmacher, M. Hinder, E. Brief, O. Gassmann, D. Hartl, Benchmarking R&D success rates of leading pharmaceutical companies: an empirical analysis of FDA approvals (2006–2022), *Drug Discov. Today*, **30** (2025), 104291. <https://doi.org/10.1016/j.drudis.2025.104291>
3. N. Q. K. Le, T. X. Tran, P. A. Nguyen, T. T. Ho, V. N. Nguyen, Recent progress in machine learning approaches for predicting carcinogenicity in drug development, *Expert Opin. Drug Metab. Toxicol.*, **20** (2024), 621–628. <https://doi.org/10.1080/17425255.2024.2356162>
4. M. Korn, C. Ehrt, F. Ruggiu, M. Gastreich, M. Rarey, Navigating large chemical spaces in early-phase drug discovery, *Curr. Opin. Struct. Biol.*, **80** (2023), 102578. <https://doi.org/10.1016/j.sbi.2023.102578>
5. Q. Yao, Z. Chen, Y. Cao, H. Hu, Enhancing drug-target interaction prediction with graph representation learning and knowledge-based regularization, *Front. Bioinf.*, **5** (2025), 1649337. <https://doi.org/10.3389/fbinf.2025.1649337>
6. A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, et al., ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.*, **40** (2012), D1100–D1107. <https://doi.org/10.1093/nar/gkr777>
7. UniProt Consortium, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Res.*, **47** (2019), D506–D515. <https://doi.org/10.1093/nar/gky1049>
8. UniProt Consortium, UniProt: a hub for protein information, *Nucleic Acids Res.*, **43** (2015), D204–D212. <https://doi.org/10.1093/nar/gku989>
9. M. Talo, S. Bozdog, Top-DTI: integrating topological deep learning and large language models for drug–target interaction prediction, *Bioinformatics*, **41** (2025), i133–i141. <https://doi.org/10.1093/bioinformatics/btaf183>
10. F. Gentile, V. Agrawal, M. Hsing, A. T. Ton, F. Ban, U. Norinder, et al., Deep docking: a deep learning platform for augmentation of structure based drug discovery, *ACS Cent. Sci.*, **6** (2020), 939–949. <https://doi.org/10.1021/acscentsci.0c00229>
11. W. Shi, H. Yang, L. Xie, X. X. Yin, Y. Zhang, A review of machine learning-based methods for predicting drug–target interactions, *Health Inf. Sci. Syst.*, **12** (2024), 30. <https://doi.org/10.1007/s13755-024-00287-6>
12. Y. Ding, J. Tang, F. Guo, Q. Zou, Identification of drug–target interactions via multiple kernel-based triple collaborative matrix factorization, *Briefings Bioinf.*, **23** (2022), bbab582. <https://doi.org/10.1093/bib/bbab582>
13. F. Chen, Z. Zhao, Z. Ren, K. Lu, Y. Yu, W. Wang, Prediction of drug target interaction based on under sampling strategy and random forest algorithm, *PLoS One*, **20** (2025), e0318420. <https://doi.org/10.1371/journal.pone.0318420>
14. R. Rodríguez-Pérez, J. Bajorath, Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery, *J. Comput.-Aided Mol. Des.*, **36** (2022), 355–362. <https://doi.org/10.1007/s10822-022-00442-9>
15. D. Song, Y. Chen, Q. Min, Q. Sun, K. Ye, C. Zhou, et al., Similarity-based machine learning support vector machine predictor of drug-drug interactions with improved accuracies, *J. Clin. Pharm. Ther.*, **44** (2019), 268–275. <https://doi.org/10.1111/jcpt.12786>

16. Y. C. Wang, Z. X. Yang, Y. Wang, N. Y. Deng, Computationally probing drug-protein interactions via support vector machine, *Lett. Drug Des. Discov.*, **7** (2010), 370–378. <https://doi.org/10.2174/157018010791163433>
17. Q. Ning, Y. Wang, Y. Zhao, J. Sun, L. Jiang, K. Wang, et al., DMHGNN: Double multi-view heterogeneous graph neural network framework for drug-target interaction prediction, *Artif. Intell. Med.*, **159** (2025), 103023. <https://doi.org/10.1016/j.artmed.2024.103023>
18. Y. Jing, D. Zhang, L. Li, H2GnnDTI: hierarchical heterogeneous graph neural networks for drug–target interaction prediction, *Bioinformatics*, **41** (2025), btaf117. <https://doi.org/10.1093/bioinformatics/btaf117>
19. O. Zhang, H. Lin, X. Zhang, X. Wang, Z. Wu, Q. Ye, et al., Graph neural networks in modern AI-aided drug discovery, *Chem. Rev.*, **125** (2025), 10001–10103. <https://doi.org/10.1021/acs.chemrev.5c00461>
20. A. G. Vrahatis, K. Lazaros, S. Kotsiantis, Graph attention networks: a comprehensive review of methods and applications, *Future Internet*, **16** (2024), 318. <https://doi.org/10.3390/fi16090318>
21. M. K. Subramanian, J. Spencer, Deep neural networks for in silico prediction of drug-target interactions: Advancements and future directions, *Front. Emerg. Artif. Intell. Mach. Learn.*, **2** (2025), 1–12.
22. Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, *IEEE Trans. Neural Networks Learn. Syst.*, **33** (2021), 6999–7019. <https://doi.org/10.1109/TNNLS.2021.3084827>
23. A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.*, **25** (2012).
24. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 770–778. <https://doi.org/10.1109/CVPR.2016.90>
25. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, *Adv. Neural Inf. Process. Syst.*, **30** (2017).
26. D. Guo, D. Yang, H. Zhang, J. Song, P. Wang, Q. Zhu, et al., DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning, *Nature*, **645** (2025), 633–638. <https://doi.org/10.1038/s41586-025-09422-z>
27. H. Öztürk, A. Özgür, E. Ozkirimli, DeepDTA: deep drug–target binding affinity prediction, *Bioinformatics*, **34** (2018), 821–829. <https://doi.org/10.1093/bioinformatics/bty593>
28. I. Lee, J. Keum, H. Nam, DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences, *PLoS Comput. Biol.*, **15** (2019), e1007129. <https://doi.org/10.1371/journal.pcbi.1007129>
29. K. Wang, R. Zhou, Y. Li, M. Li, DeepDTAF: a deep learning method to predict protein–ligand binding affinity, *Briefings Bioinf.*, **22** (2021), bbab072. <https://doi.org/10.1093/bib/bbab072>
30. K. Huang, C. Xiao, L. M. Glass, J. Sun, MolTrans: molecular interaction transformer for drug–target interaction prediction, *Bioinformatics*, **37** (2021), 830–836. <https://doi.org/10.1093/bioinformatics/btaa880>

31. Q. Zhao, H. Zhao, K. Zheng, J. Wang, HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism, *Bioinformatics*, **38** (2022), 655–662. <https://doi.org/10.1093/bioinformatics/btab715>
32. P. Bai, F. Miljković, B. John, H. Lu, Interpretable bilinear attention network with domain adaptation improves drug–target prediction, *Nat. Mach. Intell.*, **5** (2023), 126–136. <https://doi.org/10.1038/s42256-022-00605-1>
33. N. Q. K. Le, Predicting emerging drug interactions using GNNs, *Nat. Comput. Sci.*, **3** (2023), 1007–1008. <https://doi.org/10.1038/s43588-023-00555-7>
34. T. M. Nguyen, T. Nguyen, T. M. Le, T. Tran, Gefa: early fusion approach in drug–target affinity prediction, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **19** (2021), 718–728. <https://doi.org/10.1109/TCBB.2021.3094217>
35. L. Peng, X. Liu, L. Yang, L. Liu, Z. Bai, M. Chen, et al., BINDTI: a bi-directional intention network for drug–target interaction identification based on attention mechanisms, *IEEE J. Biomed. Health Inf.*, **29** (2024), 1602–1612. <https://doi.org/10.1109/JBHI.2024.3375025>
36. A. Xiong, Z. Luo, Y. Xia, Q. Zou, L. Wei, Z. Zhang, et al., An interpretable geometric graph neural network for enhancing the generalizability of drug–target interaction prediction, *BMC Biol.*, **23** (2025), 350. <https://doi.org/10.1186/s12915-025-02456-9>
37. J. Wei, Y. Zhu, L. Zhuo, Y. Liu, X. Fu, F. Li, Efficient deep model ensemble framework for drug–target interaction prediction, *J. Phys. Chem. Lett.*, **15** (2024), 7681–7693. <https://doi.org/10.1021/acs.jpclett.4c01509>
38. Z. Zhou, Q. Liao, J. Wei, L. Zhuo, X. Wu, X. Fu, et al., Revisiting drug–protein interaction prediction: a novel global-local perspective, *Bioinformatics*, **40** (2024), btae271. <https://doi.org/10.1093/bioinformatics/btae271>
39. R. Zhang, M. Yang, Q. Zhao, J. Wang, BiBLDR: Bidirectional behavior learning for drug repositioning, *IEEE J. Biomed. Health Inf.*, (2025). <https://doi.org/10.1109/JBHI.2025.3628673>
40. M. K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology, *Nucleic Acids Res.*, **44** (2016), D1045–D1053. <https://doi.org/10.1093/nar/gkv1072>
41. H. Liu, J. Sun, J. Guan, J. Zheng, S. Zhou, Improving compound–protein interaction prediction by building up highly credible negative samples, *Bioinformatics*, **31** (2015), i221–i229. <https://doi.org/10.1093/bioinformatics/btv256>
42. C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.*, **20** (1995), 273–297. <https://doi.org/10.1007/BF00994018>
43. L. Breiman, Random forests, *Mach. Learn.*, **45** (2001), 5–32. <https://doi.org/10.1023/A:1010933404324>
44. T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, S. Venkatesh, GraphDTA: predicting drug–target binding affinity with graph neural networks, *Bioinformatics*, **37** (2021), 1140–1147. <https://doi.org/10.1093/bioinformatics/btaa921>

45. L. Chen, X. Tan, D. Wang, F. Zhong, X. Liu, T. Yang, et al., TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments, *Bioinformatics*, **36** (2020), 4406–4414. <https://doi.org/10.1093/bioinformatics/btaa524>
46. X. Zeng, W. Chen, B. Lei, CAT-DTI: cross-attention and Transformer network with domain adaptation for drug-target interaction prediction, *BMC Bioinf.*, **25** (2024), 141. <https://doi.org/10.1186/s12859-024-05753-2>
47. Y. Zhang, J. Li, S. Zhao, H. Tan, H. Sun, Y. Xiong, et al., Prediction of drug-target interactions based on hypergraph neural networks with multimodal feature fusion, *IEEE J. Biomed. Health Inf.*, (2025). <https://doi.org/10.1109/JBHI.2025.3625158>
48. M. Gao, D. Zhang, Y. Chen, Y. Zhang, Z. Wang, X. Wang, et al., Graphormerdti: a graph transformer-based approach for drug-target interaction prediction, *Comput. Biol. Med.*, **173** (2024), 108339. <https://doi.org/10.1016/j.compbiomed.2024.108339>
49. L. Lebreton, O. Curet, S. Gueddari, F. Mazouz, S. Bernard, C. Burstein, et al., Selective and Potent Monoamine Oxidase Type B Inhibitors: 2-Substituted 5-Aryltetrazoles Derivatives, *J. Med. Chem.*, **38** (1995), 4786–4792. <https://doi.org/10.1021/jm00024a006>
50. I. Abdullah, C. F. Chee, Y. K. Lee, S. S. R. Thunuguntla, K. S. Reddy, K. Nellore, et al., Benzimidazole derivatives as potential dual inhibitors for PARP-1 and DHODH, *Bioorg. Med. Chem.*, **23** (2015), 4669–4680. <https://doi.org/10.1016/j.bmc.2015.05.051>
51. T. Y. Tsai, M. S. Coumar, T. Hsu, H. P. Hsieh, C. H. Chien, C. T. Chen, et al., Substituted pyrrolidine-2, 4-dicarboxylic acid amides as potent dipeptidyl peptidase IV inhibitors, *Bioorg. Med. Chem. Lett.*, **16** (2006), 3268–3272. <https://doi.org/10.1016/j.bmcl.2006.03.037>
52. J. Fiorito, F. Saeed, H. Zhang, A. Staniszewski, Y. Feng, Y. I. Francis, et al., Synthesis of quinoline derivatives: discovery of a potent and selective phosphodiesterase 5 inhibitor for the treatment of Alzheimer’s disease, *Eur. J. Med. Chem.*, **60** (2013), 285–294. <https://doi.org/10.1016/j.ejmech.2012.12.009>
53. National Center for Biotechnology Information, *PubChem BioAssay: AID 1798284*, 2026. Available from: <https://pubchem.ncbi.nlm.nih.gov/bioassay/1798284>.
54. E. Lo Piparo, H. Scheib, N. Frei, G. Williamson, M. Grigorov, C. J. Chou, Flavonoids for controlling starch digestion: structural requirements for inhibiting human α -amylase, *J. Med. Chem.*, **51** (2008), 3555–3561. <https://doi.org/10.1021/jm800115x>
55. M. I. Davis, J. P. Hunt, S. Herrgard, P. Ciceri, L. M. Wodicka, G. Pallares, et al., Comprehensive analysis of kinase inhibitor selectivity, *Nat. Biotechnol.*, **29** (2011), 1046–1051. <https://doi.org/10.1038/nbt.1990>
56. W. C. Kang, J. McAuley, Self-attentive sequential recommendation, in *2018 IEEE International Conference on Data Mining (ICDM)*, (2018), 197–206. <https://doi.org/10.1109/ICDM.2018.00035>
57. F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, et al., BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer, in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, (2019), 1441–1450. <https://doi.org/10.1145/3357384.3357895>

58. S. Yang, W. Ma, P. Sun, Q. Ai, Y. Liu, M. Cai, et al., Sequential recommendation with latent relations based on large language model, in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, (2024), 335–344. <https://doi.org/10.1145/3626772.3657762>
59. W. Zhao, Y. Yu, G. Liu, Y. Liang, D. Xu, X. Feng, et al., MSI-DTI: predicting drug-target interaction based on multi-source information and multi-head self-attention, *Briefings Bioinf.*, **25** (2024), bbae238. <https://doi.org/10.1093/bib/bbae238>
60. Y. Meng, Y. Wang, J. Xu, C. Lu, X. Tang, T. Peng, et al., Drug repositioning based on weighted local information augmented graph neural network, *Briefings Bioinf.*, **25** (2024), bbad431. <https://doi.org/10.1093/bib/bbad431>
61. S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, et al., PubChem 2023 update, *Nucleic Acids Res.*, **51** (2023), D1373–D1380. <https://doi.org/10.1093/nar/gkac956>
62. X. Chen, N. N. Guan, Y. Z. Sun, J. Q. Li, J. Qu, MicroRNA-small molecule association identification: from experimental results to computational models, *Briefings Bioinf.*, **21** (2018), 47–61. <https://doi.org/10.1093/bib/bby098>
63. Z. Zhou, L. Zhuo, X. Fu, J. Lv, Q. Zou, R. Qi, Joint masking and self-supervised strategies for inferring small molecule-miRNA associations, *Mol. Ther. Nucleic Acids*, **35** (2024), 102103. <https://doi.org/10.1016/j.omtn.2023.102103>
64. Z. Zhou, Z. Du, X. Jiang, L. Zhuo, Y. Xu, X. Fu, et al., GAM-MDR: probing miRNA-drug resistance using a graph autoencoder based on random path masking, *Briefings Funct. Genomics*, **23** (2024), 475–483. <https://doi.org/10.1093/bfgp/ela005>



AIMS Press

©2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)