



Article



SAPDTA: A Novel Protein Segment Capture Strategy for Drug-Target Affinity Prediction

Guanqiu Qi ^{1,*}, Zihao Fang ², Stanley Tang ³ and Jeffery Wang ⁴

¹ Computer Information Systems Department, State University of New York at Buffalo State, Buffalo, NY 14222, USA

² College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

³ Palo Alto High School, Palo Alto, CA 94301, USA

⁴ The Harker School, San Jose, CA 95129, USA

* Correspondence: qiq@buffalostate.edu

How To Cite: Qi, G.; Fang, Z.; Tang, S.; et al. SAPDTA: A Novel Protein Segment Capture Strategy for Drug-Target Affinity Prediction. *Sensors and AI* **2026**, *2*(1), 33–49. <https://doi.org/10.53941/sai.2026.100003>

Received: 3 March 2026

Revised: 25 May 2026

Accepted: 17 June 2026

Published: 16 July 2026

Abstract: Predicting drug-target affinity (DTA) is becoming increasingly vital in the field of drug discovery. Currently, many methods focus solely on the overall encoding of proteins, overlooking the abundant information contained within protein peptides. Therefore, this paper proposes a novel protein segment capture strategy for drug-target affinity prediction (SAPDTA), which is designed to extract local protein features through a local block capture approach. This strategy supports adaptive segmentation of amino acid chains, enabling more flexible extraction of protein structure information at different levels. A hybrid dual-network bilinear interaction module is proposed to address the challenge of protein feature extraction at various scales. Moreover, bilinear interaction blocks are employed to combine and process the chemical properties of drugs with the biological characteristics of their targets. SAPDTA's performance is assessed using two publicly accessible DTA datasets (Davis and KIBA). According to the experimental results, SAPDTA demonstrates competitive performance compared to existing models across all evaluation metrics. Furthermore, visualization results on the ToxCast dataset highlight the model's sensitivity to complex drug structures, revealing its capability to understand underlying structure-function relationships.

Keywords: deep learning; DTA prediction; feature extraction; feature fusion

1. Introduction

Drug discovery is a lengthy, expensive, and highly uncertain process, often requiring over a decade of research and development and investments amounting to billions of dollars [1]. However, it is important to note that 90% of drugs that enter clinical trials fail to receive approval from the Food and Drug Administration (FDA) for market release [2]. In recent decades, rapid advancements in computational technology have enhanced drug design in experimental settings and sped up the overall drug development process [3]. However, drug discovery and development still face significant challenges. Among these challenges, accurately predicting drug-target affinity (DTA) is a crucial aspect of drug discovery [4,5]. This prediction, a key step in the drug design process, involves estimating the binding strength between drug molecules and their biological targets, such as proteins. The structural and functional characteristics of these biological targets enable them to bind to drug molecules, and this binding ability is often closely related to pharmacological activity. Accurately predicting the affinity between drugs and biological targets is therefore essential for understanding drug-target mechanisms, guiding drug design optimization, and anticipating both therapeutic effects and potential side effects [6].

Early approaches for predicting drug-target interactions mainly included ligand-based screening, structure-based screening, and experimental methods. Ligand-based screening relies on information from known small molecules (ligands) that bind to the target protein, without directly using the protein's three-dimensional structural information [7]. Its core principle is that chemically similar molecules tend to exhibit similar biological activities.



Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Techniques such as molecular comparison, pharmacophore identification, and Quantitative Structure-Activity Relationship (QSAR) models are commonly used under this approach to predict drug-target interactions [8]. Structure-based screening leverages the three-dimensional structure of the target protein. It aims to predict the binding affinity and activity of molecules by simulating their interactions with the protein. Common methods include molecular docking, molecular dynamics simulations, and virtual screening [9]. However, all three approaches—ligand-based, structure-based, and experimental—are time-consuming and resource-intensive [10]. This creates a strong demand for more efficient methods capable of evaluating a large number of potential drug-target interactions. In response, deep learning-based approaches have recently emerged as a promising and rapidly growing area in drug discovery [11].

In recent years, the rapid advancement of computational technology and bioinformatics has highlighted the significant potential of machine learning across various fields, including medical image segmentation, bioinformatics, and drug discovery [12,13]. Traditional machine learning methods—such as decision trees, random forests, support vector machines—as well as deep learning techniques [14,15], have been widely applied to model DTA. Compared to earlier ligand-based, structure-based, and experimental approaches, deep learning models offer the advantage of automatically learning high-level feature representations, whereas traditional methods often require manual feature engineering. This capability allows deep learning models to capture complex, potentially non-linear relationships between drugs and targets, improving prediction accuracy. Additionally, deep learning can effectively integrate multi-modal data, including drug molecular structures, protein sequences and structural information, and diverse biological features [16]. By providing a more comprehensive representation of drug-target interactions, these approaches enhance predictive performance. As a result, machine learning and deep learning have become essential tools in modern drug development.

Most machine learning and deep learning approaches directly encode protein sequences for input into subsequent network layers [17]. Representing the entire protein as a one-dimensional sequence can result in the loss of intrinsic structural and functional information of the protein [18]. Proteins are typically composed of multiple functional domains, structural domains, and subunits, and this information is critical for DTA prediction [19]. Simplifying proteins to a one-dimensional sequence fails to fully capture this intrinsic information [20]. This may weaken the model's predictive power and functional relevance in the learned latent space. Simultaneously, effectively processing and integrating information from diverse sources also presents a challenges. For instance, the chemical information about drugs and the biological information about targets represent two entirely distinct data types. Effectively integrating these two types of information is key to improving prediction accuracy. To address the aforementioned challenges, a novel protein segment capture strategy for drug-target affinity prediction (SAPDTA) is proposed, in which local protein features are extracted through a local block capture approach. A hybrid dual-network bilinear interaction module is proposed to address the extraction of protein features at different scales. The dual networks in this context are mainstream deep learning models, GNN (Graph Neural Network) and CNN (Convolutional Neural Network), which can automatically learn and extract complex patterns without requiring complex feature engineering. Additionally, a bilinear interaction block is employed to combine and process the chemical properties of drugs with the biological features of their targets.

This paper has three main contributions as follows.

- A novel protein block-capturing strategy is proposed. By applying a local block-capturing method to segment protein sequences, the approach improves the extraction and interpretation of local protein features.
- A novel Temporal Hierarchical Fusion Network (THF-Net) is proposed to improve protein sequence feature extraction for DTA prediction. The network incorporates a Bi-LSTM (Bidirectional Long Short-Term Memory) module to model global sequential dependencies within protein sequences. In addition, a hierarchical attention mechanism is introduced to capture multi-scale sequence features, where local attention emphasizes short-range dependencies among protein fragments, and global attention integrates contextual information across the entire sequence. Through this hierarchical fusion strategy, THF-Net effectively learns both local and global protein sequence representations, thereby enhancing the accuracy of DTA prediction.
- A bilinear interaction block is introduced to model the interactions between drug and target input channels through a bilinear attention mechanism. This design enables the model to capture subtle interaction patterns and binding characteristics between drugs and targets. By effectively learning these complex interaction features, the model achieves more accurate predictions of drug-target binding affinities and interaction trends.

The remainder of this paper is organized as follows. Section 2 offers a review of the existing literature on deep learning approaches for DTA prediction. Section 3 elaborates on the three modules of the proposed model: the protein block capturing strategy, the hybrid dual-network bilinear interaction module, and the bilinear interaction block. Section 4 presents the experimental data, setup, and evaluation of the model on two benchmark datasets,

Davis [21] and KIBA [22], along with the corresponding results and analysis. Section 5 provides the conclusion and outlines potential directions for future research.

2. Related Work

Deep learning has found widespread utility across diverse bioinformatics domains [23]. In DTA prediction, deep learning-based architectures excel at capturing intricate, non-linear relationships from vast datasets by automatically extracting high-level feature representations. This capability is vital for DTA tasks, where interactions span complex biological processes and diverse molecular characteristics influenced by numerous hidden variables. Based on the underlying network architectures used for feature extraction, these approaches generally fall into two categories: sequence-based methods modeling both drugs and targets, and graph-based methods explicitly mapping drug molecular structures.

2.1. Sequence-Based Methods for Drugs and Targets

As a pioneer in deep learning-based DTA prediction, DeepDTA [24] utilized one-dimensional representations: Simplified Molecular Input Line Entry System (SMILES) strings for drug chemical structures and amino acid sequences for proteins. In this architecture, both SMILES strings and protein sequences are encoded as integer vectors. DeepDTA employs two identical CNNs—each featuring three 1D convolutional layers with pooling—to independently extract latent feature embeddings for drugs and proteins. The resulting vector pairs are then concatenated and fed into fully connected layers for final affinity prediction. Expanding on this work, Öztürk et al. introduced WideDTA [25], an extension that enriches sequence representations with higher-order, text-based features [24]. Specifically, it represents drugs via Ligand Maximum Common Substructures (LMCS) [26]) and proteins through Protein Domains and Motifs (PDM) from PROSITE [27]. By integrating these two additional text sources with the original sequence embeddings, WideDTA leverages four distinct representations to boost prediction accuracy while incorporating attention mechanisms to improve model interpretability. Building on these baseline architectures, subsequent studies integrated recurrent and transformer-based components to capture sequential dependencies. Zheng et al. [28] proposed DTI-RCNN (drug-target interaction—region-based convolutional neural network), an efficient hybrid model that embeds recurrent neural networks into a CNN framework. Moving toward attention-driven paradigms, Chen et al. introduced TransformerCPI (compound-protein interaction) [29], which modifies the traditional transformer encoder and final linear layer to enhance interpretability during prediction. More recently, Zhu et al. [30] developed a DTA prediction method driven by interactive learning and an automatic encoder mechanism. Parallel efforts have focused on optimizing feature extraction from raw sequences. Lee et al. [31] developed DeepConv-DTI (deep convolutional neural network for drug-target interaction), which processes raw amino acid sequences and drug fingerprint strings. By utilizing 1D-CNNs with varied kernel sizes, DeepConv-DTI effectively captures local residual conformational patterns in proteins, mitigating the performance risks of fixed, sub-optimal kernel choices. Finally, Yang et al. developed ML-DTI (mutual learning mechanism for interpretable drug-target interaction prediction) [32], which introduces a mutual information mechanism to bridge the gap between drug and protein encoders. This approach explicitly captures the interactive relationships between the two encoders, ultimately improving the identification of novel drug-target interactions.

2.2. Graph-Based Methods for Drug Molecules

Transitioning toward structural representations, Tsubaki et al. [33] developed a hybrid model integrating a Graph Neural Network (GNN) for drug molecular graphs, a CNN for protein sequences, and an attention mechanism to fuse their representations. Leveraging 2D structural features has proven highly effective; for instance, Rifaioğlu et al. [34] introduced DEEPScreen, a large-scale system that inputs compounds as 2D structural images and employs CNNs to extract complex visual features for high-precision prediction. Meanwhile, Abbasi et al. [35] proposed DeepCDA (deep cross-domain compound-protein affinity prediction), combining convolutional and Long Short-Term Memory (LSTM) layers to encode both local features and global temporal patterns for cross-domain affinity prediction. To fully exploit molecular topology, several frameworks have moved exclusively toward graph-based architectures. Nguyen et al. [36] introduced GraphDTA (graph drug-target affinity), which represents drugs as undirected graphs defined by feature maps and adjacency matrices. Building on this paradigm, Cheng et al. [37] implemented an end-to-end method utilizing graph attention networks coupled with multi-head self-attention. Expanding graph representations to both interaction partners, Jiang et al. [38] proposed DGraphDTA (dynamic graph drug-target affinity), which models drug structures and protein contact maps as separate graphs, applying parallel GNNs to extract their respective embeddings. Recent innovations have focused on optimizing these graphs and filtering cross-domain features. Nguyen et al. [39] developed GEFA (graph early fusion affinity), a graph-in-

graph neural network with an attention mechanism designed to capture conformational changes in target proteins induced by binding effects. Similarly, TDGraphDTA (transformer and diffusion on drug graphs in drug-target affinity prediction), proposed by Zhu et al. [40], optimizes drug molecular graphs to preserve node-association information, extracting rich structural features while enhancing model interpretability. Finally, Liao et al. [41] introduced GSAML, an interpretable drug-target binding affinity prediction model integrating self-attention mechanism, mutual information and GNNs to construct target-protein and drug representations. Notably, their model incorporates a mutual information objective to filter out redundant feature noise while retaining the core interactive signals critical for binding affinity prediction.

3. The Proposed Solution

The overall architecture of SAPDTA is illustrated in Figure 1. As an advanced deep learning framework designed for DTA prediction, SAPDTA executes its pipeline across three primary stages.

- Stage 1—Drug Feature Extraction: The chemical structure of the drug is converted from its SMILES string into a molecular graph representation. This topological format allows graph-based deep learning algorithms to effectively analyze and learn intricate molecular patterns.
- Stage 2—Target Protein Representation: The target protein sequence is adaptively segmented into multi-scale peptide fragments using a protein block capturing strategy. These fragments are then processed by the THF-Net for feature extraction. Within THF-Net, a Bi-LSTM module captures global sequential dependencies, while a hierarchical attention mechanism extracts features across multiple granularities. Specifically, a local attention mechanism resolves short-range dependencies within individual protein fragments, while a global attention mechanism aggregates information across the entire sequence to yield a comprehensive protein embedding.
- Stage 3—Bilinear Interaction and Affinity Prediction: The drug's graph representation and the protein sequence's feature vector are fed into a bilinear interaction module, which models complex cross-modal associations between the input features. These integrated representations are finally processed through a series of fully connected layers to output the predicted DTA value.

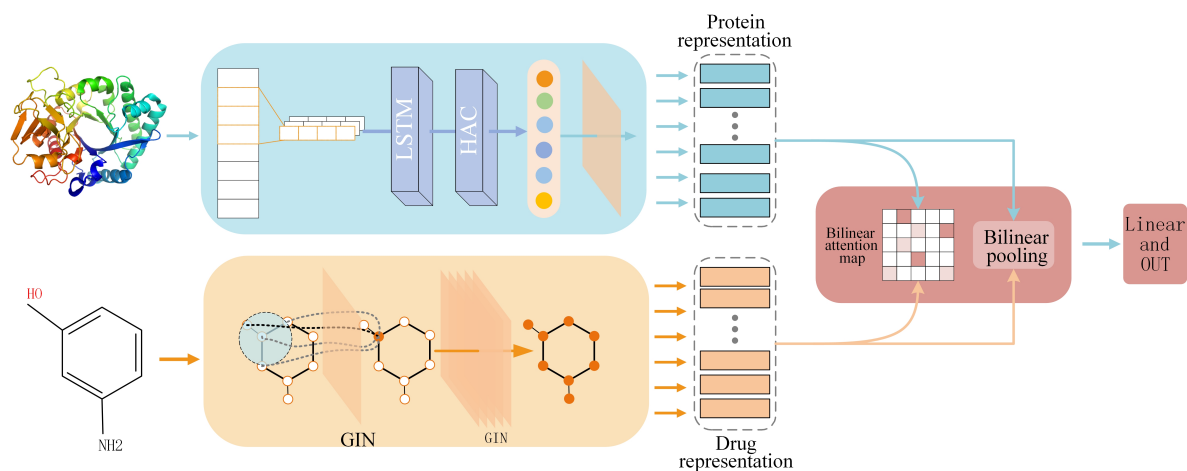


Figure 1. The overall architecture of SAPDTA. It is designed to process drug structures and protein sequences simultaneously. For the drug branch, structural features are extracted through multiple graph convolutional layers, followed by a global sum pooling layer to generate a fixed-size representation. Concurrently, the protein sequence is partitioned into continuous fragments; each segment is embedded and processed through a hybrid dual-network bilinear interaction module. These processed protein fragments are then averaged and pooled to yield a fixed-length representation of the target. Finally, the extracted drug and protein feature embeddings are fed into a central bilinear interaction module for cross-modal fusion.

3.1. Protein Block Capturing Strategy

Proteins, as critical molecules within living organisms, often have functions closely associated with specific protein segments, known as peptides. For instance, certain peptide segments in the body serve as enzymes, catalyzing specific biochemical reactions. Meanwhile, some peptides act as signaling molecules, facilitating information transfer between cells. Moreover, the precise structure of peptides often determines their interaction mode and binding affinity with other molecules, such as proteins, DNA (deoxyribonucleic acid), RNA (ribonucleic acid), or small molecules. The structure and function of protein sequences often exhibit a high degree of complexity, with

various regions within them often carrying unique biological attributes. Slicing these sequences into shorter segments allows for a more precise exploration of the protein's details, further revealing its local functions and structural characteristics. By subdividing it into protein fragments, one can more accurately capture the distinct properties of these segments. This strategy enables an in-depth examination of the attributes of different regions within the protein sequence and their crucial roles in biological processes. Finally, through the analysis and comparison of these segment features, a deeper exploration of the protein's functional diversity and structural variability can be achieved. The structure of the protein block capturing strategy is depicted in Figure 2.

To better capture the hierarchical structural information within protein sequences, this paper proposes a dynamic windowing strategy that enables adaptive importance weights for different fragment scales during training. Unlike traditional methods that use fixed-size sliding windows for feature extraction, the proposed strategy allows the model to extract multi-scale local and global features by dynamically controlling the granularity of sequence segmentation. The specific steps are given as follows. (1) Protein fragments are first generated offline using several fixed window sizes. (2) These fragments are then encoded and fed into the model simultaneously. (3) During training, the attention/gating mechanism learns which fragment scales contribute more to the affinity prediction task. (4) The learnable parameters associated with fragment selection and aggregation are optimized via backpropagation, while the actual window-size values remain fixed and discrete.

Given an input protein sequence represented as a tensor $X \in \mathbb{R}^{N \times T}$, where N is the batch size and T is the sequence length, the model applies multiple window sizes $W_s \in \{W_1, W_2, \dots, W_S\}$ to segment each sequence. At scale s , the sequence is divided into $T - W_s + 1$ fragments:

$$w_s^{(n,i)} \in \mathbb{R}^{W_s}, \quad i = 1, 2, \dots, T - W_s + 1$$

Each fragment is then transformed into a feature vector via a shared embedding layer:

$$f_s^{(n,i)} = \mathbf{W}_{\text{embed}} w_s^{(n,i)} + \mathbf{b}_{\text{embed}}$$

Feature representations from different scales are later fused to support multi-granular learning of structural properties.

The proposed method does not directly optimize discrete window sizes through gradient backpropagation. Instead, multiple predefined fixed window sizes are used to generate protein fragments offline. These fragments are processed simultaneously by the model, while an attention or gating mechanism learns adaptive importance weights for different fragment scales during training. Backpropagation updates only the weighting and representation-learning components, not the discrete window-size values themselves.

To support this process, this paper designs a composite loss function that jointly optimizes prediction accuracy, feature granularity, local consistency, and window-size regularity:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{basic}} + \lambda_2 \mathcal{L}_{\text{multi-scale}} + \lambda_3 \mathcal{L}_{\text{local}} + \lambda_4 \mathcal{L}_{\text{reg}}$$

The components are defined as follows:

- (1) Basic loss $\mathcal{L}_{\text{basic}}$, which measures the prediction error, e.g., mean squared error (MSE) for regression tasks:

$$\mathcal{L}_{\text{basic}} = \frac{1}{N} \sum_{n=1}^N \left(y^{(n)} - \hat{y}^{(n)} \right)^2$$

- (2) Multi-scale loss $\mathcal{L}_{\text{multi-scale}}$, which evaluates the contribution of each scale to the prediction:

$$\mathcal{L}_{\text{multi-scale}} = \sum_{s=1}^S \alpha_s \cdot \left[\frac{1}{N} \sum_{n=1}^N \sum_{i=1}^{T-W_s+1} \left(f_s^{(n,i)} - y^{(n)} \right)^2 \right]$$

where α_s is the weighting coefficient for scale s .

- (3) Local structure consistency loss $\mathcal{L}_{\text{local}}$, which ensures that similar fragments yield similar feature representations:

$$\mathcal{L}_{\text{local}} = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^{T-W+1} \sum_{j=i+1}^{T-W+1} S(w^{(n,i)}, w^{(n,j)})$$

where $S(\cdot, \cdot)$ is a similarity function, such as cosine similarity.

- (4) Regularization loss $\mathcal{L}_{\text{reg}}$, which constrains excessive variation in window sizes:

$$\mathcal{L}_{\text{reg}} = \sum_{s=1}^S (W_s - \overline{W})^2$$

where $\overline{W}$ is a predefined reference window size.

This composite loss function enables the model to jointly learn accurate predictions and adaptive segmentation strategies. Particularly, the $\mathcal{L}_{\text{multi-scale}}$ and $\mathcal{L}_{\text{reg}}$ terms cooperate to ensure that the model effectively balances the influence of different window granularities while avoiding overfitting to any specific scale. The effectiveness and generalizability of this design will be further validated in the experimental section.

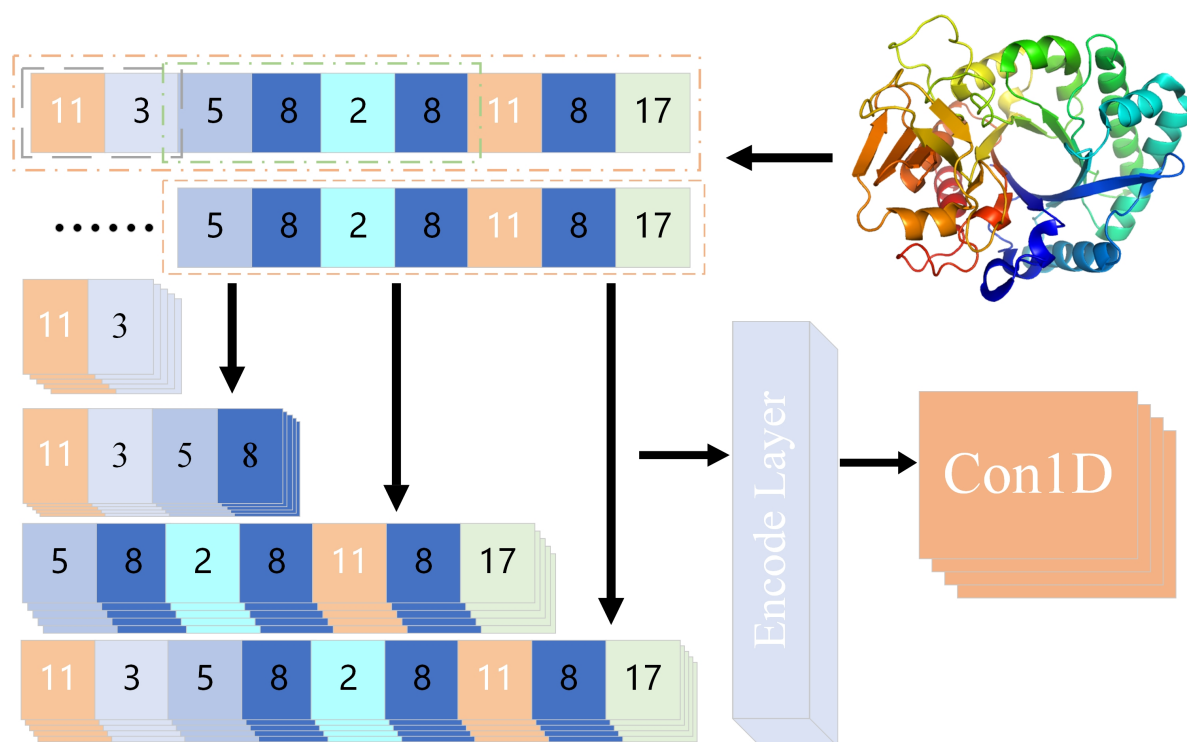


Figure 2. Protein Block Capturing Strategy module.

3.2. Temporal Hierarchical Fusion Network Module

In this study, this paper proposes the THF-Net to enhance the feature extraction capabilities of protein sequences, especially in the context of DTA prediction. Figure 3 illustrates the protein encoding pipeline, where each amino acid is first mapped to a dense embedding and then processed through bidirectional LSTM layers to capture contextual sequence information. An attention mechanism highlights the most informative residues, and the outputs are aggregated into a fixed-length protein representation for downstream drug-target affinity prediction. This approach integrates both local and global sequence features in a unified representation.

As the input sequence, the protein sequence is represented as individual amino acids, shown as letters M (Methionine), E (Glutamic Acid/Glutamate), T (Threonine), L (Leucine). Each amino acid is first passed through an encoding layer, which maps it into a dense embedding vector (the green circles shown in Figure 3). This embedding captures semantic and physicochemical properties of each amino acid.

The module integrates two key components—LSTM module and hierarchical attention mechanism—which, through their collaborative operation, improve the comprehensive understanding of multi-dimensional protein features (referring to interactions across multiple feature channels).

First, the LSTM module handles the temporal information within protein sequences, capturing both long-term and short-term dependencies between protein fragments. The functionality and structure of proteins are often influenced by temporal variations in their sequences, making the capture of temporal features essential for understanding their biological activities. Unlike traditional unidirectional LSTMs, our module uses bidirectional LSTM, which learns the sequence information from both forward and backward contexts, allowing it to capture the intrinsic complexity of protein sequences more comprehensively and accurately. The bidirectional LSTM effectively captures the global temporal dependencies within the protein sequence, providing rich temporal information for subsequent feature fusion.

Second, the introduction of a hierarchical attention mechanism enables the model to extract multi-scale features.

In protein sequences, both local and global features are equally important. The local attention mechanism focuses on short-range dependencies within protein fragments, helping to capture detailed features in local regions, while the global attention mechanism aggregates information across the entire protein sequence, capturing long-range dependencies and forming a comprehensive protein representation. The hierarchical attention mechanism allows the model to simultaneously process protein features at multiple scales, enhancing its ability to capture the complexity and diversity of protein sequences.

Then, by combining the LSTM module and the hierarchical attention mechanism, the Temporal Hierarchical Fusion Network Module deeply integrates temporal features and local features, enabling the comprehensive capture of multi-dimensional protein characteristics. The outputs from the forward and backward LSTMs for each position are combined, producing a sequence of contextualized feature vectors (shown as colored circles in Figure 3). These vectors represent multi-dimensional features for each amino acid, integrating sequential and contextual information. This design enhances the model's capacity to represent protein sequences, thereby improving the accuracy of DTA predictions.

Next, the sequence of LSTM outputs is passed through an attention layer. The attention mechanism computes weights for each residue feature, highlighting the most important positions that contribute to the protein's functional behavior or drug-binding properties.

Finally, the weighted features are aggregated into a fixed-length protein representation vector (shown in the dashed box in Figure 3). This vector serves as the input to downstream modules, such as bilinear interaction networks or drug-target affinity prediction layers.

Consequently, the THF-Net module couples the strengths of LSTMs with hierarchical attention mechanisms to deliver a highly accurate, comprehensive characterization of biomolecular behavior. This dual-engine approach plays a pivotal role in optimizing both drug discovery pipelines and protein function predictions.

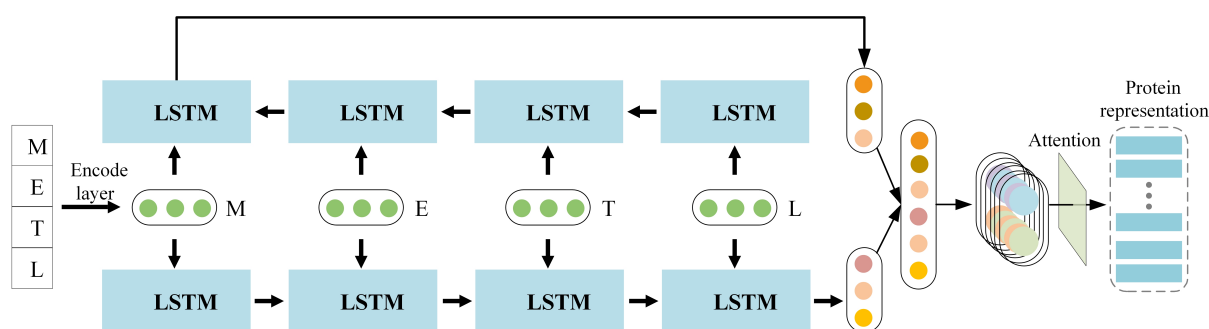


Figure 3. Temporal Hierarchical Fusion Network Module. It highlights the flow from raw sequence → embedding → bidirectional LSTM → attention → final protein representation.

3.3. Hybrid Dual-Network Bilinear Interaction Module

Adapted from the bilinear attention network, the bilinear interaction module is a specialized deep learning component engineered to fuse drug and protein features for DTA prediction, as illustrated in Figure 4. The protein branch extracts hierarchical features via multi-scale feature blocks and convolutional layers, while the drug branch encodes molecular features using a similar convolutional pipeline. The features from both branches are fused, passed through fully connected layers, and output a predicted binding affinity or interaction score. Originally designed for Visual Question Answering (VQA) tasks, the bilinear attention network [42] represents a powerful framework for multi-modal learning. By leveraging a novel bilinear attention mechanism, this architecture explicitly models pairwise local interactions across different data modalities. Since VQA and DTA prediction share a similar multi-modal problem structure, this bilinear attention network framework is modified to introduce a pairwise interaction module tailored for DTA modeling.

In the protein branch (the top one shown in Figure 4), the protein input (shown as a 3D structure) is first processed through a Multi-Scale Feature Block (MFB) to capture hierarchical features. It then passes through a sequence of convolutional layers (Conv) → ReLU → Conv → Softmax, which extract local and contextual patterns. Two separate feature streams (Feat.1 and Feat.2) are generated, representing different scales or perspectives of protein features. These features are concatenated to form a unified protein representation vector.

In the drug branch (the bottom one shown in Figure 4), the drug is represented as a molecular graph or structure (example shown is a small molecule with -OH and -NH₂ groups). It is processed through a Conv → ReLU → Conv → Softmax pipeline to extract chemical features. The output features are denoted as Feat and capture the molecular properties relevant for binding.

As the process of fusion and interaction, the drug features are fed into the protein branch's MFB (dashed arrow shown in Figure 4), allowing cross-modal interaction between drug and protein representations. After fusion, the combined representation passes through a fully connected layer (FC) → Softmax → OUT, producing the final output, which could be a binding affinity score or classification of binding vs. non-binding.

In this network, there are two multi-channel inputs, X and Y. To reduce the dimensionality of the two input channels, a bilinear attention matrix is $\mathcal{A} \in \mathbb{R}^{\rho \times \phi}$ introduced as follows:

$$\mathbf{f}'_k = (\mathbf{X}^T \mathbf{U}')^T_k \mathcal{A} (\mathbf{Y}^T \mathbf{V})_k \quad (1)$$

where $\mathbf{U}' \in \mathbb{R}^{N \times K}$, $\mathbf{V}' \in \mathbb{R}^{M \times K}$, $(\mathbf{X}^T \mathbf{U}')_k \in \mathbb{R}^\rho$, and $\mathbf{f}'_k$ represents the K-th element in the middle, with the subscript K denoting the column index.

The bilinear attention network function can be represented as:

$$\mathbf{f} = \text{BAN}(\mathbf{X}, \mathbf{Y}; \mathcal{A}) \quad (2)$$

The bilinear pooling attention distribution map is obtained by applying a softmax function to the output:

$$\mathcal{A}_g := \text{softmax}(((1 \cdot \mathbf{p}_g^T) \circ \mathbf{X}^T \mathbf{U}) \mathbf{V}^T \mathbf{Y}) \quad (3)$$

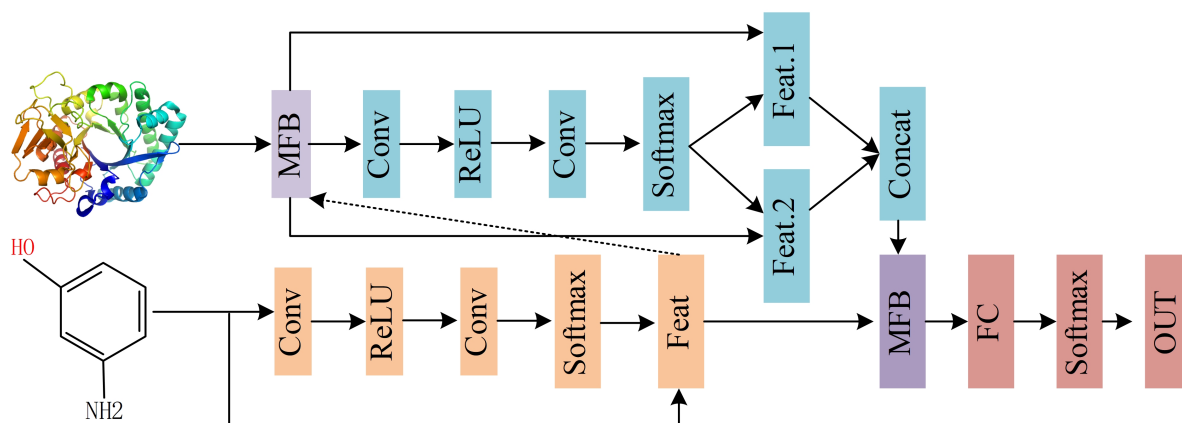


Figure 4. Hybrid Dual-Network Bilinear Interaction Module

4. Experiment

4.1. Evaluation Metrics

In order to compare with the baseline model, the performance of the model is evaluated using the Concordance Index (CI) and Mean Squared Error (MSE). The Concordance Index (CI) is a model evaluation metric proposed by Gönen and Heller [43], designed to measure the difference between model predictions and true values. CI can be used to assess the relative ranking of drug-target pairs (whether predicted affinities correctly rank drug-target pairs). In drug discovery, relative ranking is often more important than the exact affinity value. Suppose there are two drug-target pairs (d_i, t_i) and (d_j, t_j) . Let y_i and y_j be their true binding affinities, and $\hat{y}_i$ and $\hat{y}_j$ be predicted affinities. CI measures the fraction of pairs where the predicted ordering agrees with the true ordering. The formula is given as follows.

$$CI = \frac{1}{N} \sum_{i,j} h(\hat{y}_i - \hat{y}_j), \quad (4)$$

for all pairs where $y_i \neq y_j$, where

$$N = \text{total number of comparable pairs } (y_i \neq y_j), \quad (5)$$

$$h(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0.5, & \text{if } x = 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (6)$$

CI = 1 → perfect ranking, CI = 0.5 → random ranking, CI < 0.5 → worse than random (predictions often reversed).

MSE (Mean Squared Error) measures the magnitude of errors between predicted and true affinities, which is widely used in regression tasks. Unlike CI, it cares about exact predicted values, not just the ranking.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2, \quad (7)$$

where n = number of drug-target pairs, y_i = true affinity, and $\hat{y}_i$ = predicted affinity. Squared difference penalizes large errors more than small ones. Lower MSE means better predictive accuracy. MSE is useful for quantitative prediction tasks, where knowing the approximate binding strength matters. It is often combined with CI for a full picture. CI measures correct ranking. MSE measures closeness of predictions.

4.2. Dataset and Implementation Details

The proposed model was evaluated on two widely used benchmark datasets for DTA prediction: Davis [21] and KIBA [22]. In these datasets, binding affinities are quantified using dissociation constants (K_d), inhibition constants (K_i), or integrated KIBA scores. Since raw K_d and K_i values span several orders of magnitude and decrease as binding affinity increases, they are converted into logarithmic values (pK_d and pK_i) so that higher metrics reflect stronger binding profiles. Conversely, lower values in the KIBA dataset denote higher binding affinities. A statistical summary of these benchmarks is provided in Table 1. The Davis dataset comprises interactions across 442 proteins and 68 compounds, whereas the larger KIBA dataset encompasses 229 proteins and 2116 compounds.

Figure 5 shows key statistics of the Davis (left) and KIBA (right) datasets. The top row displays the distribution of drug-target affinities, the middle row shows the lengths of drug SMILES strings, and the bottom row shows the lengths of protein sequences. These histograms highlight the variability in binding values and sequence lengths, which are important considerations for drug-target affinity prediction models. (1) Top row—Affinity Distribution: For Davis, the histogram shows the number of drug-target (DT) pairs at different binding affinities. Most DT pairs are clustered around lower affinity values. For KIBA, the histogram shows the distribution of KIBA scores across DT pairs, with the majority concentrated around the central score range. (2) Middle row—SMILES Length Distribution: It displays the distribution of drug molecule lengths in characters. For Davis, most molecules have SMILES strings between ~40–90 characters. For KIBA, the distribution is broader, with some drugs having SMILES longer than 400 characters. (3) Bottom row—Protein Sequence Length Distribution: It shows the distribution of protein sequence lengths in amino acids. In Davis, sequences are mostly below 2000 characters, while KIBA contains some proteins longer than 4000 characters.

All experimental tests were conducted on a computer with the following specifications: NVIDIA GeForce RTX 4090 with 24.00 GB of VRAM (NVIDIA, Santa Clara, CA, USA), Intel Core i7-12700 CPU @ 2.10 GHz (Intel, Santa Clara, CA, USA), and 64.00 GB of RAM. This study used standard five-fold cross-validation, where the data was evenly divided into five subsets and in each iteration four subsets are used for training (training set) and the remaining subset for evaluation (test set). Detailed hyperparameter settings are provided in Table 2. Specifically, the graph neural network for extracting compounds consists of five GIN (graph isomorphism network) layers. A batch size of 512 and training was performed for 1000 epochs. The Adam optimizer with a learning rate of 0.0005 was used to update model parameters.

Table 1. Summary of the datasets Davis and KIBA.

Dataset	Number of Compounds	Number of Proteins	Number of Paired Samples
Davis	68	442	30,056
KIBA	2,111	229	118,254

Table 2. Hyperparameter configurations.

Hyperparameter	Setting
Learning rate	0.0005
Batch size	512
Epoch	1,000
Optimizer	Adam
GIN layers	5

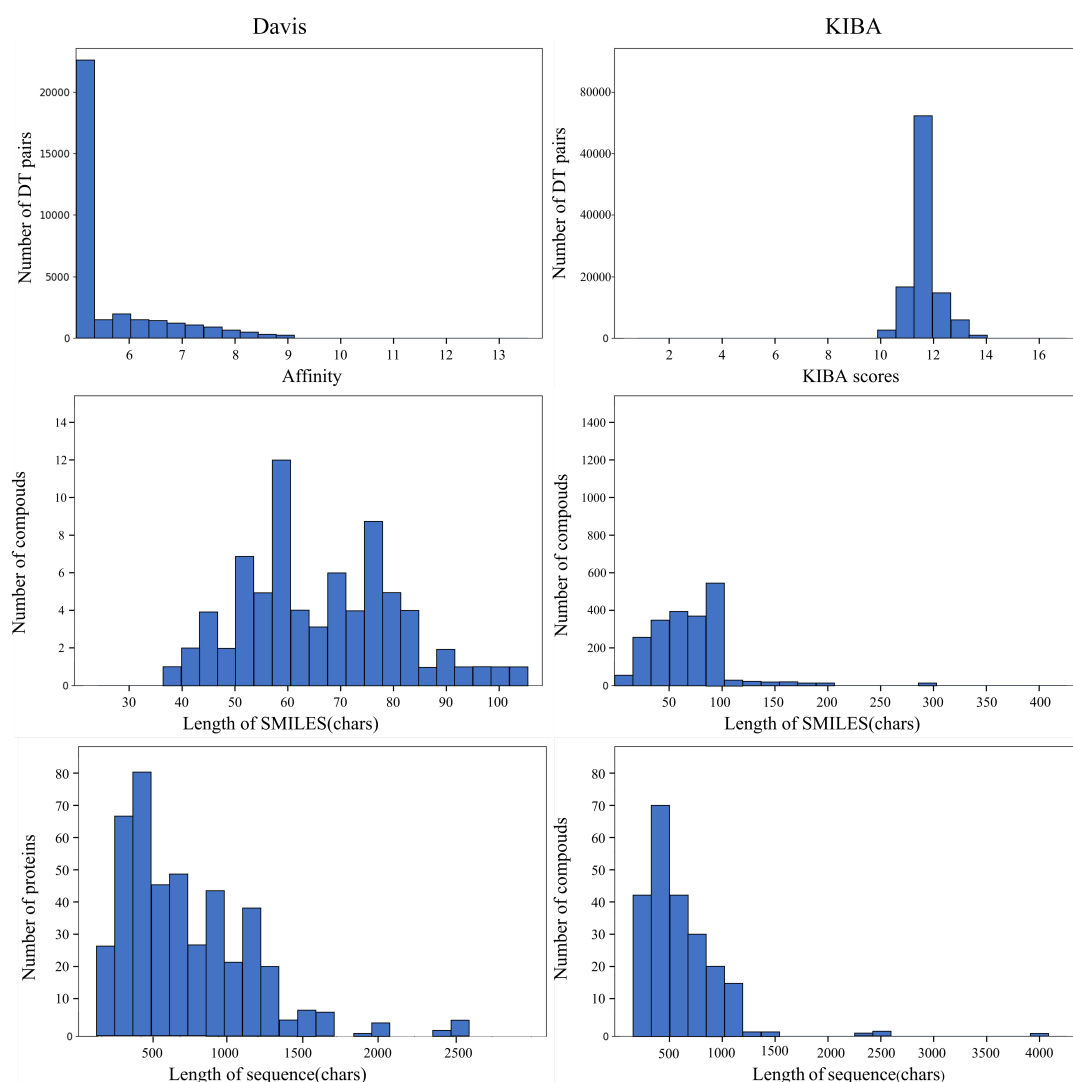


Figure 5. The distribution of the Davis and KIBA datasets. The first row shows the distribution of binding affinity values. The second row shows the distribution of SMILES string length. The third row shows the distribution of protein sequence length.

4.3. Results of Predictive Performance of Random Splitting

To evaluate the effectiveness of the proposed SAPDTA model, its predictive performance was compared with existing methods on the Davis and KIBA datasets. For regression-based evaluation, two quantitative metrics were employed: mean squared error (MSE), where lower values indicate better performance, and concordance index (CI), where higher values reflect superior predictive capability. The proposed SAPDTA model was compared to traditional models such as DeepDTA [24], WideDTA [25], MATT-DTI (multiple attention—drug-target interaction) [44], as well as state-of-the-art models (Bicomp-DTA (biological-related and compression-based drug-target affinity) [45], DoubleSG-DTA (double sequence and graph—drug-target affinity) [46], MT-DTI (molecule transformer—drug-target interaction) [47], Graph-DTA (graph drug-target affinity) [36], ELECTRA-DTA (efficiently learning an encoder that classifies token replacements accurately—drug-target affinity) [48]).

For regression tasks on the Davis and KIBA datasets, the proposed SAPDTA model was compared with several state-of-the-art approaches, including DoubleSG-DTA and Bicomp-DTA. A five-fold cross-validation strategy was adopted, and the final performance was evaluated on the test set under the same experimental settings as SAPDTA. To ensure a fair comparison, the data samples in each fold were kept consistent across all methods.

Tables 3 and 4 display the performance of the SAPDTA method compared to existing models on the Davis and KIBA datasets. In both datasets, SAPDTA demonstrates competitive performance compared to existing models across all evaluation metrics. In most cases, it achieves the highest or comparable values, highlighting its effectiveness while acknowledging that the improvements may be marginal in some datasets. Specifically, on the Davis dataset, SAPDTA improves the CI score by 0.011 and reduces the MSE by 0.011 compared with the baseline model GraphDTA. On the KIBA dataset, SAPDTA achieves an improvement of 0.01 in CI and a reduction of 0.015

in MSE relative to GraphDTA. In addition, for models that previously proposed multiple architectural variants, only the best-performing architecture on the Davis dataset is reported.

Figure 6 illustrates the relationship between real affinity and predicted values on the Davis (left plot) and KIBA (right plot) datasets. X-axis, which is labeled “Actual Labels” on both plots, represents the true ground-truth values. Y-axis is labeled as “Vertical Distance to $y = x$ ” on both plots. While labeled as “Vertical Distance”, the axis values and the diagonal line indicate that the y-axis is plotting the Predicted Values, and the green diagonal line represents the identity line ($y = x$) where the prediction perfectly equals the actual label. The more concentrated the points are on the line ($y = x$), the better the performance of the corresponding method. In the left plot (Davis dataset), the x-axis ranges from 5 to 12, and the y-axis ranges from 5 to 12. The data points (blue dots) are heavily concentrated in the lower-left region, specifically between the actual values of 5 and 9. For actual values between 5 and 7, the model shows a tendency to overestimate, as many points sit well above the green line. As the actual labels increase past 8, the data points fall predominantly below the green line, showing that the model begins to underestimate higher values. The data completely cuts off around an actual label value of 10.5, and the predicted values max out just under 10. In the right plot (KIBA dataset), the x-axis ranges from 9 to 17, and the y-axis ranges from 9 to 17. This plot appears to capture a higher range of the same dataset. The data points are much more densely packed and uniformly distributed along the green identity line compared to the left plot. The points form a distinct, upward-sloping cluster that tightly tracks the $y = x$ line from 9 all the way up to 16. While there is still variance (a spread of about 2 to 3 units above and below the line), the overall linear relationship is much stronger here, indicating better predictive accuracy across this higher numerical range. As shown in Figure 6, the samples in both the Davis and KIBA datasets are distributed approximately symmetrically around the line $y=x$, with the KIBA dataset exhibiting a denser concentration of points near this line. Based on the results presented in Tables 3 and 4, and Figure 6, SAPDTA demonstrates competitive performance across all evaluation metrics, achieving either the best or comparable results. These findings indicate that SAPDTA provides more reliable and accurate predictive performance than other DTA models on the Davis and KIBA datasets.

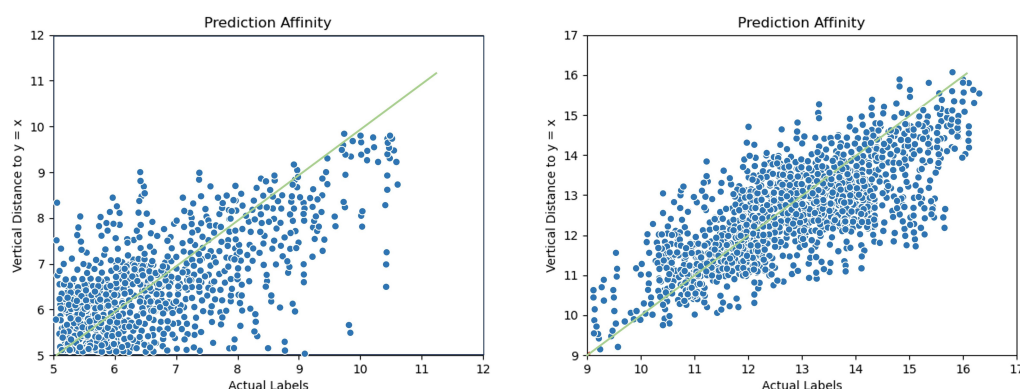


Figure 6. Drug-target affinity prediction results obtained by SAPDTA on the Davis (**left**) and KIBA (**right**) datasets

Table 3. The performance of SAPDTA and baseline models on the Davis dataset.

Model	Protein Rep.	Compound Rep.	CI	MSE
WideDTA	1D + PDM rep	1D + LMCS	0.886	0.262
DeepDTA	1D	1D	0.878	0.261
MATT-DTI	1D	L.E + E.L + SA + FC + 1DCNN	0.890	0.227
Graph-DTA	1D	Graph(GAT-GCN)	0.881	0.245
Graph-DTA	1D	Graph(GIN)	0.893	0.230
ELECTRA-DTA	-	-	0.897	0.238
Bicomp-DTA	BiComp + FC	L.E + E.L + 1CNN + SepCNN	0.903	0.237
MT-DTI	-	-	0.887	0.245
DoubleSG-DTA	SENet	GIN + SENet	0.902	0.219
SAPDTA (proposed)	Splitting + LSTM + 1D	Graph	0.904	0.219

PDM: protein domains and motifs; rep: replication-associated protein; LMCS: ligand maximum common substructures; MATT-DTI: multiple attention—drug-target interaction; L.E: label encoding; E.L: embedding layer; SA: self-attention block; FC: fully-connected block; 1DCNN: one-dimensional convolutional neural network; GAT-GCN: graph attention network—graph convolutional network; GIN: graph isomorphism network; ELECTRA-DTA: efficiently learning an encoder that classifies token replacements accurately—drug-target affinity; Bicomp-DTA: biological-related and compression-based drug-target affinity; SepCNN: depthwise separable convolutional neural network; MT-DTI: molecule transformer—drug-target interaction; DoubleSG-DTA: double sequence and graph—drug-target affinity; SENet: squeeze-and-excitation network; LSTM: long short-term memory.

Table 4. The performance of SAPDTA and baseline models on the KIBA dataset.

Model	Protein Rep.	Compound Rep.	CI	MSE
WideDTA	1D + PDM rep	1D + LMCS	0.875	0.179
DeepDTA	1D	1D	0.863	0.194
MATT-DTI	1D	L.E + L.ESA + FC + 1DCNN	0.889	0.150
Graph-DTA	1D	Graph(GAT-GCN)	0.891	0.139
Graph-DTA	1D	Graph(GIN)	0.882	0.147
ELECTRA-DTA	-	-	0.889	0.162
Bicomp-DTA	BiComp+FC	L.E + E.L + 1CNN + SepCNN	0.891	0.167
DoubleSG-DTA	SENet	GIN + SENet	0.891	0.167
SAPDTA (proposed)	Splitting + LSTM + 1D	Graph	0.901	0.124

PDM: protein domains and motifs; rep: replication-associated protein; LMCS: ligand maximum common substructures; MATT-DTI: multiple attention—drug-target interaction; L.E: label encoding; E.L: embedding layer; SA: self-attention block; FC: fully-connected block; 1DCNN: one-dimensional convolutional neural network; GAT-GCN: graph attention network—graph convolutional network; GIN: graph isomorphism network; ELECTRA-DTA: efficiently learning an encoder that classifies token replacements accurately—drug-target affinity; Bicomp-DTA: biological-related and compression-based drug-target affinity; SepCNN: depthwise separable convolutional neural network; MT-DTI: molecule transformer—drug-target interaction; DoubleSG-DTA: double sequence and graph—drug-target affinity; SENet: squeeze-and-excitation network; LSTM: long short-term memory.

4.4. Ablation Study

In order to further substantiate the efficacy of the protein segment capture strategy designed within the framework of our methodology, as well as the key components including the hybrid dual-network bilinear interaction module and bilinear interaction blocks, this section conducts a dissection study. The results are shown in Tables 5 and 6.

Table 5. Objective evaluation results for the ablation study on the Davis dataset.

Model	CI	MSE
Graph-DTA	0.893	0.230
Graph-DTA + SAP	0.895	0.227
Graph-DTA + SAP+ HAC	0.899	0.223
Graph-DTA + SAP + HAC + BIB	0.904	0.219

Table 6. Objective evaluation results for the ablation study on the KIBA dataset.

Model	CI	MSE
Graph-DTA	0.891	0.139
Graph-DTA + SAP	0.895	0.135
Graph-DTA + SAP + HAC	0.898	0.129
Graph-DTA + SAP + HAC + BIB	0.901	0.124

In the current experimental setup, the standard Graph-DTA is employed as the baseline model, subsequently incrementally incorporating distinct components to assess their respective impacts. Specifically, this paper primarily compares the performance of the following four models:

Graph-DTA: This model exclusively employs the Graph-DTA framework for drug-target interaction (DTA) prediction, serving as the baseline model.

Graph-DTA + Protein Segment Capture Strategy(SAP): In this variant, the protein segment capture strategy is introduced into the Graph-DTA model for segmentation purposes.

Graph-DTA + Protein Segment Capture Strategy(SAP) + Temporal Hierarchical Fusion Network Module(HAC): Building upon the aforementioned models, this iteration incorporates the hybrid dual-network bilinear interaction module. Graph-DTA + Protein Segment Capture Strategy(SAP) + Temporal Hierarchical Fusion Network Module(HAC) + Bilinear Interaction Blocks(BIB): This comprehensive model encompasses the entirety of the proposed components, culminating in a holistic approach to DTA prediction.

Therefore, the comparisons conducted between Graph-DTA and Graph-DTA + Protein Segment Capture Strategy serve to substantiate the effectiveness of employing the protein segment capture strategy in drug-target interaction (DTA) prediction. Furthermore, the juxtaposition of Graph-DTA + Protein Segment Capture Strategy and Graph-DTA + Protein Segment Capture Strategy + Temporal Hierarchical Fusion Network Module aims to validate the efficacy of the Temporal Hierarchical Fusion Network Module. Finally, the comparison between

Graph-DTA + Protein Segment Capture Strategy + Temporal Hierarchical Fusion Network Module and Graph-DTA + Protein Segment Capture Strategy + Temporal Hierarchical Fusion Network Module + Bilinear Interaction Blocks is intended to demonstrate the effectiveness of the Bilinear Interaction Blocks.

4.5. Visualization of Toxic Substructure Attention

Deep learning models are often considered “black-box” approaches because it is difficult to interpret which features contribute to their predictions. This lack of interpretability limits its application in biomedical domains. To explore the attention mechanism of the SAPDTA model, this paper introduces Grad-AAM (Gradient-weighted Affinity Activation Mapping), a visualization tool that highlights atom-level importance, and applies it to the ToxCast dataset, which contains various toxicological properties.

The model is trained on ToxCast following the same procedure described earlier. Grad-AAM is then used to generate attention maps over molecular structures to assess whether SAPDTA can identify key toxic substructures. This study focuses on four representative toxic functional groups containing the core elements O, C, S, and Br: carboxyl groups (O-based), alkynes and haloalkyls (C-based), thioethers (S-based), and brominated groups (Br-based). These functional groups are frequently associated with mutagenic or carcinogenic effects and can interact with biomolecules such as DNA, proteins, and lipids, leading to the formation of harmful compounds or toxic metabolites.

As shown in Figure 7, SAPDTA visualizes toxic structural regions in 2D molecular representations, with color intensity reflecting the attention scores, ranging from 0 to 1. The model effectively identifies toxic moieties and demonstrates selective focus on more toxic substructures when multiple are present within a molecule. Although toxicity prediction and affinity prediction serve different objectives—safety evaluation vs. binding capability—they are structurally related. Chemical features that enhance target binding can also contribute to toxicity through nonspecific interactions with biomolecules. This “structural duality” suggests that local structural features learned during affinity prediction may be transferable to toxicity modeling. Visual interpretation of the model’s structural focus across tasks helps reveal shared mechanisms and representational capabilities.

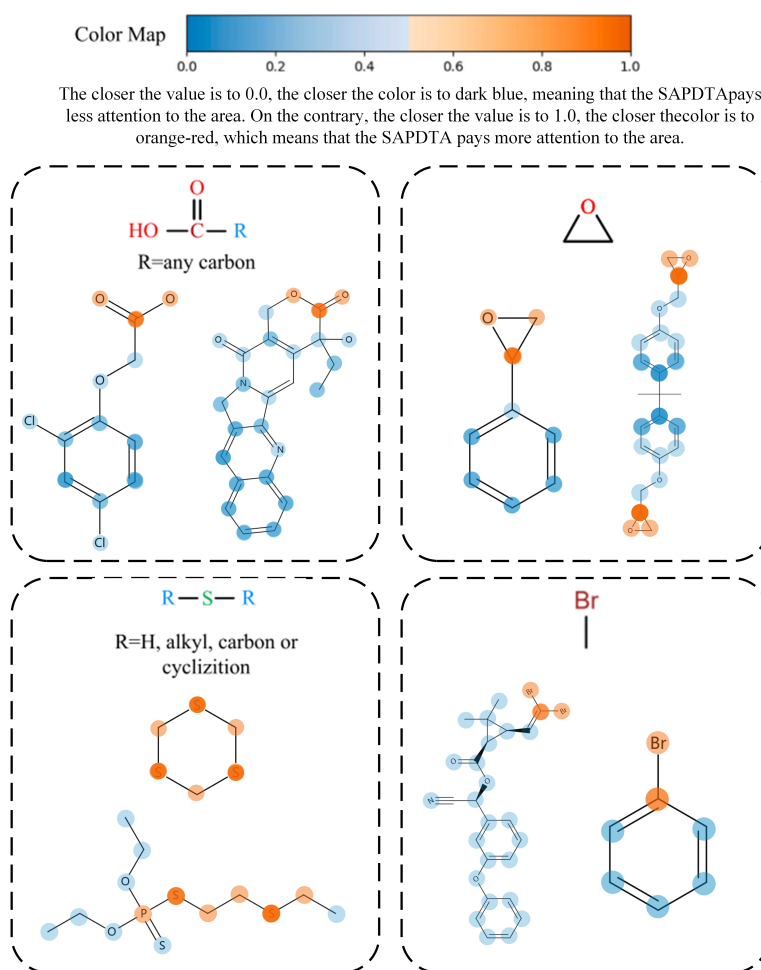


Figure 7. Toxicity structure attention visualization by SAPDTA on the ToxCast

This section visualizes test samples to further illustrate the effectiveness of the proposed method. As shown in Figure 7, the blue region represents the predicted area, while the yellow point indicates the actual binding site. It can be observed that the protein binding site generally falls within the predicted blue region. Although some prediction errors remain, the true binding points are still largely encompassed by this region. However, as with other methods, this visualization primarily reflects computational attention and does not directly convey clear biomedical interpretation. To address this limitation, future work will further investigate the functional regions of drug molecules interacting with proteins to improve the biomedical interpretability of the model.

4.6. Visualization of Predicted Protein-Drug Binding Structures

Figure 8 displays the structural models of four distinct protein-ligand complexes, labeled (a)–(d), generated by the proposed method. Each panel features the full protein structure on the left as a light purple ribbon diagram with the predicted binding site highlighted. For enhanced clarity, the right side of each panel provides a zoomed-in view of the binding pocket. Here, ligands are shown as sticks (cyan carbon atoms), while key interacting residues are highlighted in yellow and teal.

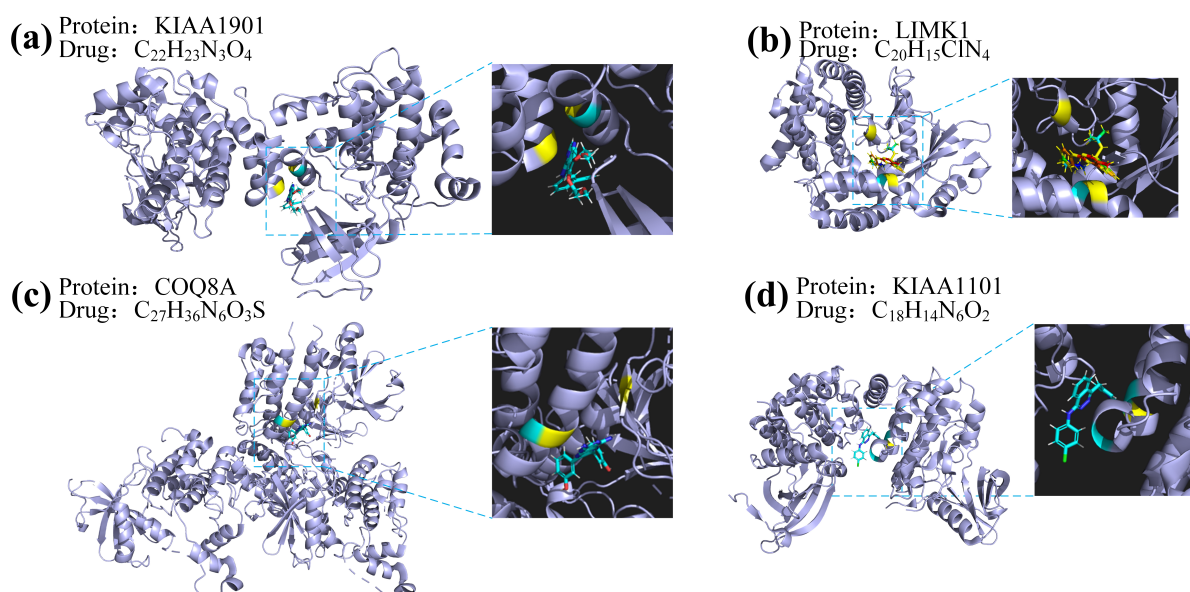


Figure 8. Predicted Protein-Drug Binding Structures for Four Targets. (a) protein KIAA1901, drug $C_{22}H_{23}N_3O_4$, (b) protein LIMK1, drug $C_{20}H_{15}ClN_4$, (c) protein COQ8A, drug $C_{27}H_{36}N_6O_3S$, (d) protein KIAA1101, drug $C_{18}H_{14}N_6O_2$

- (a) Protein KIAA1901—Drug $C_{22}H_{23}N_3O_4$: The KIAA1901 protein adopts a multi-domain folded structure with a well-defined central binding cavity. The ligand is positioned within a pocket formed by several α -helices and loop regions. The zoomed-in view reveals close contact between the ligand and surrounding residues, indicating stabilizing interactions such as hydrogen bonds and hydrophobic contacts. The pocket appears moderately deep, enclosing the ligand in a resemblance-buried conformation.
- (b) Protein LIMK1—Drug $C_{20}H_{15}ClN_4$: The LIMK1 structure displays a compact kinase-like fold. The ligand binds within a narrow cleft, likely corresponding to the active or ATP-binding site. The enlarged inset highlights the ligand nestled between helices and β -strands, with nearby residues forming a tight interaction network. The presence of a chlorine atom in the ligand may contribute to hydrophobic or halogen bonding interactions within the pocket.
- (c) Protein COQ8A—Drug $C_{27}H_{36}N_6O_3S$: COQ8A is shown as a larger, multi-subunit or multi-domain assembly with a central binding region. The ligand is located near the core of the structure, suggesting functional relevance to enzymatic activity. The zoomed panel reveals a relatively spacious binding pocket accommodating the larger ligand, with multiple contact points involving polar and nonpolar residues. The sulfur-containing group in the ligand may participate in specific interactions within the pocket.
- (d) Protein KIAA1101—Drug $C_{18}H_{14}N_6O_2$: KIAA1101 exhibits a bilobal architecture, with the ligand bound at the interface of two domains. The binding pocket appears more exposed compared to the other proteins. In the close-up view, the ligand is oriented along a helical segment, forming interactions primarily with residues lining the groove. The binding mode suggests a combination of π - π stacking and hydrogen bonding interactions.

Across all panels, Figure 8 emphasizes the diversity of protein folds and ligand-binding environments. Despite differences in size and topology, each protein accommodates its respective ligand within a distinct pocket, highlighting the specificity of molecular recognition. The consistent visualization style allows easy comparison of binding modes and structural contexts. The proposed method demonstrates good general applicability.

5. Conclusions

This paper presents a novel deep learning-based framework, SAPDTA, for predicting DTA. In this approach, the chemical structures of drugs are first converted into molecular graph representations using SMILES strings, and multiple graph convolutional layers are employed to extract meaningful drug features. At the same time, target protein sequences are partitioned into several segments through a “Protein Block Capture Strategy”, which facilitates a more detailed exploration of structural and functional characteristics of proteins. The extracted protein segments are then processed by the Temporal Hierarchical Fusion Network Module, where LSTM networks capture temporal dependencies within protein sequences, while hierarchical attention mechanisms enable multi-scale feature learning for a richer representation of protein properties. Finally, drug and protein features are integrated through a Bilinear Interaction Module based on a bilinear attention network. This module explicitly models local interactions between drugs and proteins, leading to more accurate prediction of their affinity relationships.

To validate the effectiveness of the proposed SAPDTA model, performance evaluations were conducted on the benchmark Davis and KIBA datasets against several established baseline models. The experimental results demonstrate that SAPDTA achieves competitive—and in many metrics, superior—performance in predicting DTA. These findings confirm that the proposed framework serves as a robust predictive tool capable of significantly reducing both the time and financial costs traditionally associated with the computational drug discovery and screening pipeline.

Author Contributions

G.Q.: conceptualization, methodology, software, writing—original draft preparation, reviewing and editing, supervision, investigation; Z.F.: software, data curation, writing—original draft preparation, investigation; S.T.: data curation, visualization, writing—reviewing and editing; J.W.: data curation, validation, writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The datasets used in this study are publicly available. The Davis dataset is accessible at <https://github.com/dingyan20/Davis-Dataset-for-DTA-Prediction>, and the KIBA dataset is available at <https://github.com/hkmztrk/DeepDTA>.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

1. Riemer, A.; Freund, V. Generative artificial intelligence in pharmaceutical drug development: A systematic review of time and cost efficiency across discovery, preclinical, and clinical phases. *Intell. Pharm.* **2026**, *4*, 145–158.
2. Newman, D.J.; Cragg, G.M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803.

3. Lin, X.; Li, X.; Lin, X. A review on applications of computational methods in drug screening and design. *Molecules* **2020**, *25*, 1375.
4. Soh, J.; Park, S.; Lee, H. HIDTI: integration of heterogeneous information to predict drug-target interactions. *Sci. Rep.* **2022**, *12*, 3793.
5. Zhu, Z.; Ding, Y.; Qi, G.; et al. Drug-target affinity prediction using rotary encoding and information retention mechanisms. *Eng. Appl. Artif. Intell.* **2025**, *147*, 110239.
6. Ferreyra, S.; Dutra, I.; Galeano, A.; et al. Stereochemistry-Aware Drug-Target Affinity Prediction. *bioRxiv* **2026**. Available online: <https://www.biorxiv.org/content/early/2026/05/18/2026.05.14.725200.full.pdf> (accessed on 31 May 2026).
7. Zheng, M. Ligand-Based Virtual Screening. In *Artificial Intelligence for Drug Design*; Li, H., Zheng, M., Zhu, F., et al., Eds.; Springer Nature: Singapore, 2026; pp. 745–759.
8. Ding, H.; Takigawa, I.; Mamitsuka, H.; et al. Similarity-based machine learning methods for predicting drug–target interactions: a brief review. *Briefings Bioinform.* **2014**, *15*, 734–747.
9. Dolui, S.; Mondal, A.; Biswas, K.; et al. Virtual screening and molecular dynamics simulations reveal insights into drug repurposing strategies against New Delhi metallo- β -lactamase. *Lett. Drug Des. & Discov.* **2025**, *22*, 100156.
10. Ezzat, A.; Wu, M.; Li, X.L.; et al. Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. *Briefings Bioinform.* **2019**, *20*, 1337–1357.
11. Rifaioğlu, A.S.; Atas, H.; Martin, M.J.; et al. Recent applications of deep learning and machine intelligence on *in silico* drug discovery: methods, tools and databases. *Briefings Bioinform.* **2019**, *20*, 1878–1912.
12. Bhat, A.R.; Ahmed, S. Artificial intelligence (AI) in drug design and discovery: A comprehensive review. *In Silico Res. Biomed.* **2025**, *1*, 100049. <https://doi.org/10.1016/j.ins.2025.100049>.
13. Agarwal, B.; Gaware, S.; More, N.; et al. A review exploring the translational perspective of artificial intelligence in drug discovery and formulation development. *Ann. Pharm. Fr.* **2026**, *84*, 559–594. <https://doi.org/10.1016/j.pharma.2026.01.007>.
14. Chu, Y.; Kaushik, A.C.; Wang, X.; et al. DTI-CDF: a cascade deep forest model towards the prediction of drug-target interactions based on hybrid features. *Briefings Bioinform.* **2021**, *22*, 451–462.
15. Zhu, Z.; Zheng, X.; Qi, G.; et al. Drug-target binding affinity prediction model based on multi-scale diffusion and interactive learning. *Expert Syst. Appl.* **2024**, *255*, 124647.
16. Ren, Y.; Bai, H.; Wang, J.; et al. Deep Learning-Enabled Multi-Omics Integration: A New Frontier in Precise Drug Target Discovery. *Biology* **2026**, *15*, 410.
17. Pitarch, B.; Pazos, F. Deep Learning Approaches for the Prediction of Protein Functional Sites. *Molecules* **2025**, *30*, 214. <https://doi.org/10.3390/molecules30020214>.
18. Alanazi, W.; Meng, D.; Pollastri, G. Advancements in one-dimensional protein structure prediction using machine learning and deep learning. *Comput. Struct. Biotechnol. J.* **2025**, *27*, 1416–1430. <https://doi.org/10.1016/j.csbj.2025.04.005>.
19. Yin, T.; Chen, Y.; Wang, Y.; et al. Protein structure prediction powered by artificial intelligence: from biochemical foundations to practical applications. *Front. Mol. Biosci.* **2026**, *13*, 1767821. <https://doi.org/10.3389/fmolb.2026.1767821>.
20. Cui, J.; Jiang, Y.; Wang, J.; et al. Deep learning on protein language model embeddings unlocks accurate prediction of protein solubility. *Front. Microbiol.* **2026**, *17*, 1716930. <https://doi.org/10.3389/fmicb.2026.1716930>.
21. Davis, M.I.; Hunt, J.P.; Herrgard, S.; et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2011**, *29*, 1046–1051.
22. Tang, J.; Szwajda, A.; Shakyawar, S.; et al. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J. Chem. Inf. Model.* **2014**, *54*, 735–743.
23. Bengio, S.; Wallach, H.M.; Larochelle, H.; et al. *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2018.
24. Öztürk, H.; Özgür, A.; Ozkirimli, E. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* **2018**, *34*, i821–i829.
25. Öztürk, H.; Ozkirimli, E.; Özgür, A. WideDTA: prediction of drug-target binding affinity. *arXiv* **2019**, arXiv:1902.04166.
26. Woźniak, M.; Wołos, A.; Modrzyk, U.; et al. Linguistic measures of chemical diversity and the “keywords” of molecular collections. *Sci. Rep.* **2018**, *8*, 7598.
27. Sigrist, C.J.; Cerutti, L.; De Castro, E.; et al. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.* **2010**, *38*, D161–D166.
28. Zheng, X.; He, S.; Song, X.; et al. DTI-RCNN: new efficient hybrid neural network model to predict drug–target interactions. In *Proceedings of the Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks*, Rhodes, Greece, 4–7 October 2018; pp. 104–114.
29. Chen, L.; Tan, X.; Wang, D.; et al. TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics* **2020**, *36*, 4406–4414.
30. Zhu, Z.; Yao, Z.; Qi, G.; et al. Associative learning mechanism for drug-target interaction prediction. *CAAI Trans. Intell. Technol.* **2023**, *8*, 1558–1577.
31. Lee, I.; Keum, J.; Nam, H. DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS Comput. Biol.* **2019**, *15*, e1007129.
32. Yang, Z.; Zhong, W.; Zhao, L.; et al. ML-DTI: mutual learning mechanism for interpretable drug–target interaction

- prediction. *J. Phys. Chem. Lett.* **2021**, *12*, 4247–4261.
33. 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.
 34. Rifaioğlu, A.S.; Nalbat, E.; Atalay, V.; et al. DEEPScreen: high performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations. *Chem. Sci.* **2020**, *11*, 2531–2557.
 35. Abbasi, K.; Razzaghi, P.; Poso, A.; et al. DeepCDA: deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks. *Bioinformatics* **2020**, *36*, 4633–4642.
 36. Nguyen, T.; Le, H.; Quinn, T.P.; et al. GraphDTA: Predicting drug–target binding affinity with graph neural networks. *Bioinformatics* **2021**, *37*, 1140–1147.
 37. Cheng, Z.; Yan, C.; Wu, F.X.; et al. Drug-Target Interaction Prediction Using Multi-Head Self-Attention and Graph Attention Network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *19*, 2208–2218.
 38. Jiang, M.; Li, Z.; Zhang, S.; et al. Drug–target affinity prediction using graph neural network and contact maps. *RSC Adv.* **2020**, *10*, 20701–20712.
 39. Nguyen, T.M.; Nguyen, T.; Le, T.M.; et al. Gefa: early fusion approach in drug-target affinity prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *19*, 718–728.
 40. Zhu, Z.; Yao, Z.; Zheng, X.; et al. Drug-target affinity prediction method based on multi-scale information interaction and graph optimization. *Comput. Biol. Med.* **2023**, *167*, 107621.
 41. Liao, J.; Chen, H.; Wei, L.; et al. GSAML-DTA: An interpretable drug-target binding affinity prediction model based on graph neural networks with self-attention mechanism and mutual information. *Comput. Biol. Med.* **2022**, *150*, 106145.
 42. Kim, J.H.; Jun, J.; Zhang, B.T. Bilinear attention networks. *Adv. Neural Inf. Process. Syst.* **2018**, *31*.
 43. Gönen, M.; Heller, G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika* **2005**, *92*, 965–970.
 44. Zeng, Y.; Chen, X.; Luo, Y.; et al. Deep drug-target binding affinity prediction with multiple attention blocks. *Briefings Bioinform.* **2021**, *22*, bbab117.
 45. Kalematis, M.; Zamani Emani, M.; Koohi, S. BiComp-DTA: Drug-target binding affinity prediction through complementary biological-related and compression-based featurization approach. *PLOS Comput. Biol.* **2023**, *19*, e1011036.
 46. Qian, Y.; Ni, W.; Xianyu, X.; et al. DoubleSG-DTA: Deep Learning for Drug Discovery: Case Study on the Non-Small Cell Lung Cancer with EGFR T 790 M Mutation. *Pharmaceutics* **2023**, *15*, 675.
 47. Shin, B.; Park, S.; Kang, K.; et al. Self-attention based molecule representation for predicting drug-target interaction. In Proceedings of the Machine Learning for Healthcare Conference, Ann Arbor, MI, USA, 8–10 August 2019; pp. 230–248.
 48. Wang, J.; Wen, N.; Wang, C.; et al. ELECTRA-DTA: a new compound-protein binding affinity prediction model based on the contextualized sequence encoding. *J. Cheminform.* **2022**, *14*, 14.