A Sequence-Graph Fusion Framework via BiMamba and Fourier-KAN for Interpretable Drug-Target Affinity Prediction

Xibo Li[1] **Lian Chen**[1] **Dingyuan Chen**[1] **Yichuan Zhao**[1] **Li Zhou**[*2] **Dongxi Li**[*2]

[1]College of Artificial Intelligence, Taiyuan University of Technology, Shanxi, China
[2]College of Computer Science and Technology, Taiyuan University of Technology, Shanxi, China
[*]Corresponding authors: zhouli@tyut.edu.cn, dxli0426@126.com

Abstract

Accurate prediction of drug-target affinity (DTA) is crucial for accelerating drug discovery. Existing methods struggle to model global long-range dependencies of sequences at linear complexity, and multi-layer perceptrons relying on fixed activation functions in graph neural networks face limitations in fitting complex non-linear topologies. We propose SGFF-DTA, a sequence-graph fusion framework based on bidirectional Mamba (BiMamba) and Fourier-Kolmogorov-Arnold Networks (Fourier-KAN). The framework employs BiMamba to capture global contextual semantics of sequences at linear complexity through bidirectional selective state space scanning. It integrates Fourier-KAN into graph neural networks to model high-order topological interactions using learnable non-linear transformations in the spectral space. To achieve cross-modal semantic alignment, we introduce pre-trained feature transfer and a cross-gated fusion module. On three benchmark datasets, SGFF-DTA significantly outperforms state-of-the-art methods in terms of mean squared error, demonstrating robust generalization capabilities under cold-start settings. Visual analysis further confirms model interpretability in accurately locating key binding sites and pharmacophores.

1 INTRODUCTION

New drug research and development is a expensive and time-consuming process. Bringing a novel therapeutic drug from discovery to market often exceeds a decade, with costs reaching billions [Wouters et al., 2020, Sarkar et al., 2023]. In this process, screening lead compounds is a critical step determining success [Li et al., 2023]. Drug-Target Affinity (DTA) determines specific interactions and reflects strength via quantitative metrics such as dissociation constant K_d or inhibition constant K_i [Pahikkala et al., 2015]. Unlike binary prediction identifying only binding status, DTA provides richer pharmacodynamic information crucial for evaluating biological activity, selectivity, and potential toxicity. Developing computational models for rapid, accurate DTA prediction is strategically significant for narrowing candidates, reducing trial-and-error costs, and accelerating drug discovery[Zhao et al., 2025, Zeng et al., 2024].

With deep learning breakthroughs, data-driven methods dominate the field via powerful automatic feature extraction, overcoming manual feature limitations. In early sequence explorations, DeepDTA [Öztürk et al., 2018] used convolutional neural networks (CNN) to extract features of drug and protein, while DeepGLSTM [Mukherjee et al., 2022] introduced long short-term memory networks for sequential temporal dependencies. ML-DTI [Yang et al., 2021] enhances bidirectional interaction and alignment between drug and target features via a mutual learning mechanism. Addressing missing topological information in sequence representations, graph-based methods emerged. GraphDTA [Nguyen et al., 2021] models drugs as graphs and uses graph neural networks (GNNs) to capture topological connections and atomic spatial features. DeepGS [Lin et al., 2020] combines GNNs and N-gram features to strengthen molecular representations. Addressing over-smoothing, MGraphDTA [Yang et al., 2022] constructs a deep multi-scale GNN architecture capturing local and global structures. However, most methods are limited to single-modal feature extraction, restricting deep analysis of complex drug-target interactions.

Given single-modal limitations, multimodal learning fusing sequence semantics and structural topologies dominates DTA prediction. MutualDTA [Yuan et al., 2025] uses pre-trained language models for high-quality drug-protein representations, establishing molecular-level associations via mutual attention. Similarly, DMFF-DTA [He et al., 2025] employs a bimodal fusion strategy combining sequence extractors and GNNs for efficient binding pattern modeling. While progressing, sequence encoders face an

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:3685–3704.

efficiency-performance trade-off. Traditional CNN or RNN struggle with long-range dependencies. While Transformers [Vaswani et al., 2017] establish global associations, they face quadratic complexity when processing long proteins. Linear-complexity State Space Models (SSM), especially Mamba [Gu and Dao, 2024], show potential comparable to Transformer in long sequence modeling [Waleffe et al., 2024, Wu et al., 2024b]. Although MambaTransDTA [Lou et al., 2025] integrates Mamba, it applies unidirectional causal Mamba solely to drug SMILES while employing simple CNNs for protein representation. This design neglects the non-causal 3D folding characteristics of macromolecules and does not use the linear complexity of Mamba to capture global protein sequence constraints.

Despite progress, DTA prediction faces dual bottlenecks. Current sequence models struggle to capture global protein long-range dependencies at linear complexity. Concurrently, multi-layer perceptrons (MLPs) relying on fixed activation functions exhibit limited expressive power [Liu et al., 2025], hindering the fitting of non-linear topological interactions in GNNs. To address these bottlenecks subject to biophysical constraints, we propose SGFF-DTA, a sequence-graph fusion framework featuring a synergistic integration strategy. It uses BiMamba for bidirectional selective protein scanning to extract non-causal global dependencies at linear complexity. Instead of employing an isolated paradigm, we introduce structural priors to map drug and protein contact graphs into a unified topological space. By constructing a composite graph with an explicit interaction node, our framework reflects the actual drug-target binding process and enables direct cross-modal topological message passing. We integrate Fourier-KAN into GNNs for the first time in DTA prediction, utilizing Fourier basis functions that mathematically align with molecular topologies and energy fluctuations to overcome the over-smoothing bottleneck. Furthermore, to address cross-modal initialization difficulties, we utilize a pre-trained feature transfer strategy injecting rich sequence semantics into graph nodes, achieving deep alignment and a warm start. Finally, we employ a cross-gated fusion module to achieve dynamic sequence-graph synergy. The key contributions are summarized as follows:

- We propose a sequence-graph fusion framework to overcome single-modal representational limitations. It integrates an interaction node, pre-trained feature transfer, and a cross-gated fusion module. This design achieves drug-target alignment and enhances representational capacity via complementary advantages.
- To address protein non-causality, we design BiMamba as the sequence encoder. Through bidirectional parallel scanning, it accurately captures long-range dependencies at linear complexity, effectively overcoming information decay and ensuring global semantic integrity.
- We integrate Fourier-KAN into GNN node embedding, message passing, and readout layers, designing four variants. They execute learnable spectral non-linear transformations to enhance expressive power.
- Experiments verify the SOTA performance and generalization capability. Visualization shows interpretability in locating key binding sites and pharmacophores.

2 METHODS

The framework of SGFF-DTA is illustrated in Figure 1. Our model adopts a hierarchical dual-pathway architecture synergizing sequence and graph features for DTA prediction. Specifically, the semantic sequence module extracts and fuses the global contextual dependencies of drug-target sequences to generate semantic representations. Complementarily, the composite graph module constructs a unified topological space to integrate structural information and capture complex non-linear drug-target interactions. Finally, the cross-gated fusion module integrates these multimodal representations, inputting them into an MLP to predict DTA.

2.1 DRUG AND TARGET REPRESENTATIONS

Model inputs consist of drug SMILES S_d and target protein amino acid sequence S_p. Following GFLearn [Huang et al., 2025], we utilize a tokenizer to process sequences. We decompose S_d into character-level tokens, ensuring two-character elements like Br are recognized as a single character, and S_p into residue-level tokens. Subsequently, we use RDKit [Landrum et al., 2013] to extract chemical bond information, constructing the drug molecular graph $G_D = (V_D, E_D)$, where node set V_D represents atoms and edge set E_D represents chemical bonds. To construct protein graph $G_P = (V_P, E_P)$, we retrieve the target UniProt ID and input it with the sequence into AlphaFold [Abramson et al., 2024, Fleming et al., 2025] to predict 3D structure. Based on structural information, we extract 3D coordinates r_i of C_β atoms (or C_α for glycine) for all protein residues to represent the spatial position of the i-th residue. To capture spatial proximity, we calculate Euclidean distances between residue pairs. Setting the distance threshold $\delta = 8$ Å, we define the residue contact graph adjacency matrix A_{ij}:

$$A_{ij} = \begin{cases} 1, & \text{if } \|r_i - r_j\|_2 < \delta \\ 0, & \text{otherwise} \end{cases}, \tag{1}$$

where $\|r_i - r_j\|_2$ denotes distance. An edge exists when the distance is < 8 Å[Fariselli et al., 2001, Marquez-Chamorro et al., 2014]. Inspired by DMFF-DTA [He et al., 2025], we develop a complementary binding site acquisition strategy to capture key interactions and resolve potential information loss from a single data source. We query the experimentally verified binding site set S_{uni} via UniProt [Ahmad et al., 2025] and use P2Rank [Kriv´ak and Hoksza, 2018] to predict

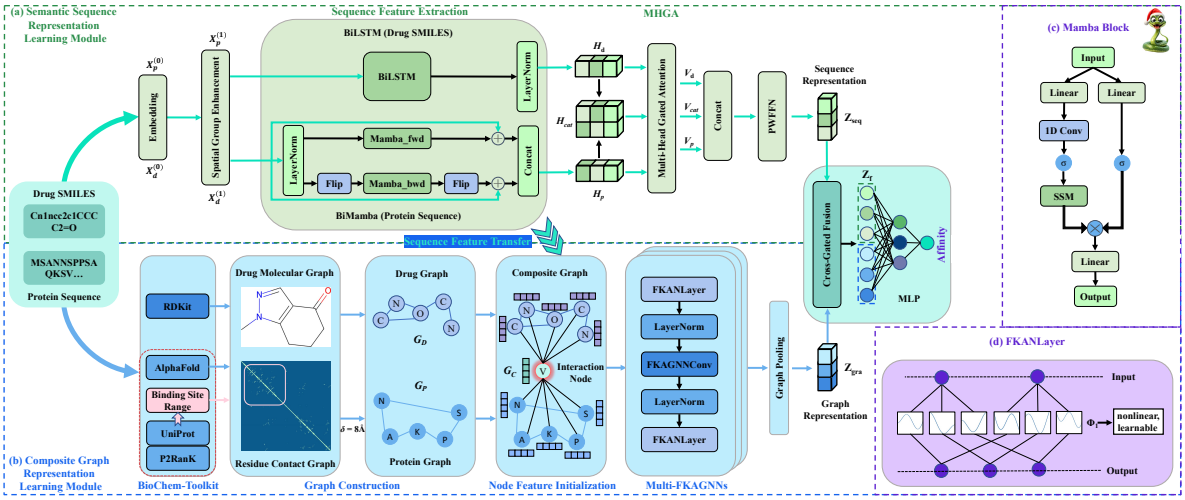

Figure 1: Overall architecture of SGFF-DTA.

potential site set S_{pred} based on the structure. The union of both sets determines the final binding site range R_{bind}:

$$R_{bind} = [\min(S_{uni} \cup S_{pred}), \max(S_{uni} \cup S_{pred})]. \quad (2)$$

We derive protein graph G_P by extracting a subgraph from the residue contact graph based on R_{bind}. Its node set V_P consists of residues within this range, and edge set E_P retains spatial contacts satisfying the distance threshold.

2.2 SEMANTIC SEQUENCE REPRESENTATION LEARNING MODULE

After the initial representation, we construct a dual-pathway semantic learning framework to extract biochemical contextual dependencies from the sequences. We map discrete token sequences into continuous dense vectors via embedding layers and linearly project to unify dimensions, generating the initial feature matrices $X_d^{(0)}$ and $X_p^{(0)}$. To suppress noise and strengthen key atom and residue features, we adapt the Spatial Group Enhancement (SGE) [Li et al., 2022] module for sequence modeling. This module divides the channel dimension into 16 groups, utilizing global statistical moments for differential semantic calibration within each group:

$$X_d^{(1)} = \mathrm{SGE}_d(X_d^{(0)}), \quad X_p^{(1)} = \mathrm{SGE}_p(X_p^{(0)}). \quad (3)$$

SGE suppresses redundancy while enhancing key features. To address the significant differences in length scales, we design a heterogeneous backbone for parallel encoding. For compact drug SMILES, we employ BiLSTM [Graves et al., 2013] to generate the hidden representation H_d:

$$H_d = \mathrm{BiLSTM}(X_d^{(1)}). \quad (4)$$

BiLSTM effectively alleviates vanishing gradients of RNN via synergistic gating, capturing local chemical linkages and short-range atomic dependencies bidirectionally. For protein sequences significantly longer than drugs, the Transformer's quadratic complexity $O(L^2)$ creates severe bottlenecks. Meanwhile, the serial computation of BiLSTM reduces efficiency on long sequences, and its fixed-parameter gating lacks content-adaptive selection, hindering long-range interaction modeling. To this end, we introduce Mamba, which discretizes continuous state space equations and adjusts transition parameters via input-dependent selection to filter noise at linear complexity. Unlike natural language sequences, which exhibit strict causal temporal dependencies, target proteins are folded macromolecular structures in which spatial interactions among amino acid residues are intrinsically non-causal. The biochemical function of a specific residue is regulated by bidirectional spatial contexts from both the N-terminus and C-terminus. Consequently, the unidirectional causal scanning paradigm of standard Mamba is sub-optimal, as it suffers from information decay when modeling long-range backward dependencies, and thus fails to capture the spatial dependencies inherent in the 3D folding of proteins. To align with this biophysical principle, we develop a BiMamba encoder to extract bidirectional global sequence dependencies with linear complexity. After applying layer normalization (LN) to input features $X_p^{(1)}$, we feed them into independently parameterized forward and backward Mamba blocks. Backward scanning flips the sequence for reverse modeling, then re-flips the output to restore order. Each path undergoes Dropout and establishes residual connections with $X_p^{(1)}$ to retain shallow features. Finally, we concatenate outputs along the feature dimension to generate the protein hidden representation H_p:

$$\begin{aligned}
\overrightarrow{H_p} &= \mathrm{Dropout}(\mathrm{Mamba}_{fwd}(\mathrm{LN}(X_p^{(1)}))) + X_p^{(1)}, \\
\overleftarrow{H_p} &= \mathrm{Dropout}(\mathrm{flip}(\mathrm{Mamba}_{bwd}(\mathrm{flip}(\mathrm{LN}(X_p^{(1)}))))) + X_p^{(1)} \\
H_p &= [\overrightarrow{H_p} \parallel \overleftarrow{H_p}].
\end{aligned} \quad (5)$$

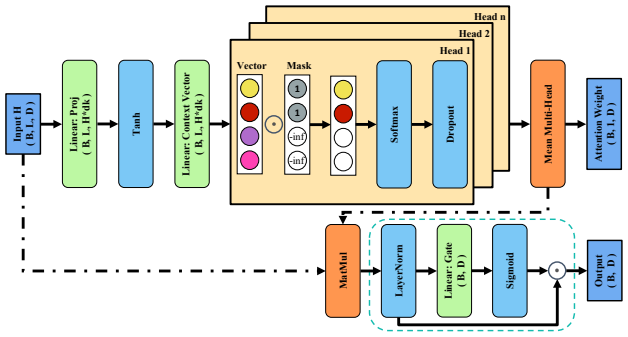

Figure 2: Architecture of the MHGA module.

Here, [‖] denotes feature concatenation, and flip represents sequence flipping. The hidden features H_d and H_p generated by these encoders contain rich contextual semantics. They serve as initial node feature sources in the subsequent graph module, transferring semantics to the graph topological space. Using H, we design a Multi-Head Gated Attention (MHGA, Figure 2) module for sequence-level aggregation, extracting key affinity features and modeling cross-modal interactions. This module projects the hidden state into a multi-head feature space via $\tanh$ activation:

$$F = \tanh(W_{proj}H + b_{proj}), \tag{6}$$

where $W_{proj} \in \mathbb{R}^{D\times(h\cdot d_k)}$ and b_{proj} are learnable parameters; h and d_k represent the number of attention heads and the feature dimension of each head. Subsequently, each head scores positional importance using an independent context vector $W_{att} \in \mathbb{R}^{h\times(h\cdot d_k)}$. A mask matrix M excludes padding. To obtain robust consensus and reduce dimensions, we average h heads for global attention weight $\bar{a}$, performing weighted aggregation to generate global feature c:

$$\bar{a} = \text{mean}_h(\text{Dropout}(\text{Softmax}(\text{mask}(W_{att}F)))), \tag{7}$$

$$c = \bar{a}H. \tag{8}$$

Finally, we introduce a gating mechanism to perform secondary screening on global features. After LN, a Sigmoid gate suppresses unimportant feature dimensions:

$$v = \text{LN}(c) \odot \text{Sigmoid}(W_{gate}\,\text{LN}(c) + b_{gate}), \tag{9}$$

where $\odot$ denotes element-wise multiplication, and $W_{gate} \in \mathbb{R}^{D\times D}$ and b_{gate} represent learnable parameters. Independently parameterized MHGA modules are applied in parallel to three perspectives: modeling internal dependencies of the drug sequence, modeling internal dependencies of the protein sequence, and modeling cross-modal interactions of the concatenated sequence, generating representation vectors v_d, v_p, and v_{cat}, respectively. After concatenation, these three vectors undergo non-linear feature fusion via a Pointwise Feed Forward Network (PWFFN) to obtain the final semantic sequence representation Z_{seq}:

$$Z_{seq} = \text{PWFFN}([v_d \parallel v_p \parallel v_{cat}]). \tag{10}$$

2.3 COMPOSITE GRAPH REPRESENTATION LEARNING MODULE

To deeply integrate drug and protein structural information and capture cross-modal interactions, we map the previously constructed molecular graph $G_D = (V_D, E_D)$ and protein graph $G_P = (V_P, E_P)$ into a unified topological space, constructing a composite graph $G_C = (V_C, E_C)$. We implement this strategy by constructing an interaction node V_i as a cross-modal information propagation hub. The node set V_C of the composite graph contains V_D, V_P, and V_i. For the edge set E_C, we retain original connection edges E_D and E_P within the subgraphs and establish bidirectional connection edges between V_i and all other nodes. This structural design establishes a cross-modal global communication mechanism, enabling the model to handle local topological details while capturing long-range spatial constraints.

To address initial feature inconsistency between drug atoms and protein residues, we employ a two-stage training strategy ensuring semantic homology. During pre-training, the model uses only the sequence modality to extract contextual features H_d and H_p. During training, we transfer H_d and H_p to composite graph nodes for feature initialization. For V_i, we define its initial feature as the average of the "[SOS]" token features from drug and protein sequences, achieving cross-modal information alignment. This pre-trained feature transfer initialization, combined with the structural inductive bias of the composite graph, significantly enhances the capability of the model to represent drug-target interactions.

To capture the non-linear topological features of the composite graph, we propose a general graph learning framework named FKAGNNs, which integrates Fourier-KAN into the GNN pipeline for the first time in DTA modeling. The core of this framework is the FKANLayer. Unlike traditional MLPs that rely on fixed activation functions, the FKANLayer is based on the Kolmogorov-Arnold representation theorem and uses Fourier series as basis functions to project input features into a spectral space for aggregation, thereby performing learnable non-linear transformations on the edges of the network. We select Fourier basis functions (harmonics) over B-splines to avoid localized computational inefficiencies during GNN message passing. Furthermore, they mathematically align with the periodic molecular topologies and local energy fluctuations inherent in drug-target binding. For an input feature x, its computation formula is:

$$\phi(x) = \sum_{k=1}^{G}(w_{c,k}\cos(kx) + w_{s,k}\sin(kx)), \tag{11}$$

where G denotes Fourier series truncation frequency; $w_{c,k}$ and $w_{s,k}$ are learnable spectral coefficients. This mechanism endows strong function fitting capabilities, achieving expressive feature transformations. Given that Graph Isomorphism

Networks (GIN) [Xu et al., 2019] theoretically possess the strongest structural discriminative power and demonstrate the best experimental performance, we focus on the improved FKAGIN model. Traditional GIN performance is limited by MLP expression bottlenecks in node embedding, message passing, and readout layers. Replacing MLP with FKANLayer in FKAGIN enables learning non-linear interactions and precisely mapping topologically isomorphic substructures to the latent feature space for better complex graph adaptability. The l-th layer FKAGIN node update is:

$$h_v^{(l)} = \phi\left((1+\epsilon^{(l)})\cdot h_v^{(l-1)} + \sum_{u\in\mathcal{N}(v)} h_u^{(l-1)}\right), \quad (12)$$

where $h_v^{(l)}$ represents the l-th layer feature vector of node v, $\mathcal{N}(v)$ is its neighbor set, and $\epsilon^{(l)}$ is a learnable parameter. After composite graph $G_C = (V_C, E_C)$ undergoes multi-layer FKAGIN message passing and aggregation, node features obtain the final graph-level structural representation Z_{gra} via Global Attention Pooling. Similarly, we apply Fourier-KAN to GraphSAGE [Hamilton et al., 2017], GCN [Kipf and Welling, 2017], and GATv2 [Brody et al., 2022]. Due to space constraints, implementation details of the variants are provided in Supplementary Material B.1.

2.4 FEATURE FUSION AND PREDICTION

After obtaining Z_{seq} and Z_{gra}, we design a Cross-Gated Fusion (CGF) module replacing simple concatenation for deep cross-modal synergy and feature refinement. This module utilizes a mutual perception gating mechanism dynamically regulating one modality via the other, retaining respective original information via residual connections. This adaptive selection mechanism effectively captures sequence-graph modality complementarity. The fusion process is:

$$\begin{aligned} Z_f = [&(Z_{seq} + Z_{seq} \odot \sigma(Z_{gra})) \\ &\| \ (Z_{gra} + Z_{gra} \odot \sigma(Z_{seq}))], \end{aligned} \quad (13)$$

where σ is the Sigmoid activation. The fused composite representation Z_f is fed into an MLP and mapped to the final drug-target affinity prediction P via a non-linear projection:

$$P = \mathrm{MLP}(Z_f). \quad (14)$$

3 RESULTS AND DISCUSSION

3.1 DATASETS AND EXPERIMENTAL SETUP

To comprehensively evaluate the predictive performance of the model, we select three widely recognized benchmark datasets in the DTA prediction: Davis [Davis et al., 2011], KIBA [Tang et al., 2014], and Metz [Metz et al., 2011]. Statistical information for each dataset is summarized in Supplementary Table S1. Davis records the kinase-inhibitor dissociation constant to measure affinity. To reduce data skewness and improve stability, original affinity values are converted into negative logarithmic values. KIBA integrates bioactivity data (including K_i, K_d, and IC_{50}) from multiple databases, using the KIBA score as a unified metric quantifying drug-target binding strength. To further verify the generalization capability of the model, we introduce Metz, adopting the negative logarithm of the inhibition constant to represent drug-target binding strength.

Our model is developed using the PyTorch framework. All experiments run on a Linux server with a 24GB VRAM NVIDIA RTX 4090D GPU. We obtain the best hyperparameters via 5-fold cross-validation and grid search on the Davis dataset, and apply them consistently to KIBA and Metz. Training employs the Adam [Kingma and Ba, 2015] optimizer. We introduce the CosineAnnealingLR to dynamically adjust learning rates. To prevent model overfitting, we set an early stopping mechanism based on validation MSE. Supplementary Table S2 provides detailed hyperparameters and default values. For all baseline models, we utilize their officially released source codes under default configurations, executing all evaluations in the identical hardware and software environment to guarantee fairness.

3.2 EVALUATION METRICS

To evaluate the multidimensional DTA prediction performance of our model, we adopt three complementary metrics: Mean Squared Error (MSE), Concordance Index (CI) [Gönen and Heller, 2005], and r_m^2 [Pratim Roy et al., 2009]. MSE widely measures the difference between predicted values (p) and true values (y). CI evaluates the model's ranking ability. It calculates the probability that the predicted relative values order of two random drug-target pairs matches their true order. The r_m^2 index rigorously validates the model's external predictive capability. r_m^2 penalizes scores when the regression line deviates from the ideal diagonal (slope 1, intercept 0), avoiding performance overestimation from linear shifts. A model shows acceptable predictive validity when $r_m^2 > 0.5$. The metric calculation formulas are:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - p_i)^2, \quad (15)$$

$$\mathrm{CI} = \frac{1}{Z}\sum_{y_i > y_j} h(p_i - p_j), \quad (16)$$

$$r_m^2 = r^2 \times (1 - \sqrt{r^2 - r_0^2}), \quad (17)$$

where N represents sample size, Z denotes data pair count, $h(x)$ is the step function (1 if $x > 0$, 0.5 if $x = 0$, 0 otherwise), and r^2 and r_0^2 represent the squared correlation coefficients with and without an intercept, respectively.

Table 1: Performance Comparison of SGFF-DTA and SOTA Methods on Davis and KIBA Datasets.

Method	Davis			KIBA		
	MSE↓	r_m^2↑	CI↑	MSE↓	r_m^2↑	CI↑
KronRLS	0.379	0.407	0.871	0.411	0.342	0.782
SimBoost	0.282	0.644	0.872	0.222	0.629	0.836
DeepDTA	0.261	0.630	0.878	0.194	0.673	0.863
DeepGS	0.252	0.686	0.882	0.193	0.684	0.860
GraphDTA_GIN	0.229	0.649	0.893	0.147	0.674	0.882
DeepCDA	0.248	0.649	0.891	0.176	0.682	0.889
MultiDTA	0.231	0.694	0.893	0.156	0.761	0.890
DeepGLSTM	0.232	0.680	0.895	0.133	0.792	0.897
ML-DTI	0.197	0.728	0.886	0.196	0.727	0.862
AttentionMGT-DTA	0.193	0.699	0.891	0.140	0.786	0.893
MGraphDTA	0.207	0.710	0.900	0.128	0.801	0.902
GLCN-DTA	0.215	0.720	0.903	0.127	0.792	0.899
LLMDTA	0.226	0.717	0.884	0.162	0.768	0.872
SMFF-DTA (2025)	0.206	0.733	0.897	0.151	0.780	0.894
DMFF-DTA (2025)	0.218	0.702	0.894	0.144	0.773	0.889
MambaTransDTA (2025)	0.191	0.737	0.894	0.173	0.722	0.871
MF-DTA (2025)	0.211	0.733	0.906	0.137	0.791	0.903
MutualDTA (2025)	0.196	0.748	0.908	0.127	0.802	**0.908**
SGFF-DTA_GCN	0.176	0.784	0.902	0.131	0.802	0.896
SGFF-DTA_GraphSAGE	0.178	0.778	0.905	0.129	0.797	0.901
SGFF-DTA_GATv2	0.175	0.782	0.908	0.128	0.810	0.899
SGFF-DTA_GIN	**0.173**	**0.789**	**0.910**	**0.125**	**0.813**	0.904

Note: ↓ denotes lower is better; ↑ denotes higher is better. Bold values indicate the best performance, and underlined values indicate the second best.

3.3 PERFORMANCE COMPARISON WITH EXISTING METHODS ON DAVIS AND KIBA

To comprehensively evaluate the performance of SGFF-DTA, we compare with 18 state-of-the-art methods on Davis and KIBA. These baselines cover broad computational strategies, including machine learning (KronRLS [Pahikkala et al., 2015], SimBoost [He et al., 2017]); sequence-based methods (DeepCDA [Abbasi et al., 2020], DeepDTA [Öztürk et al., 2018], ML-DTI [Yang et al., 2021], DeepGLSTM [Mukherjee et al., 2022]); graph-based or multimodal methods (DeepGS [Lin et al., 2020], GraphDTA [Nguyen et al., 2021], MGraphDTA [Yang et al., 2022], AttentionMGT-DTA [Wu et al., 2024a], MultiDTA [Deng et al., 2024], GLCN-DTA [Qi et al., 2024]); and cutting-edge models (LLMDTA [Tang et al., 2025], MambaTransDTA [Lou et al., 2025], SMFF-DTA [Wang et al., 2025], MF-DTA [Kang et al., 2025], DMFF-DTA [He et al., 2025], MutualDTA [Yuan et al., 2025]). To ensure a fair comparison, all experiments use 5-fold cross-validation under identical data splits. We randomly divide Davis and KIBA into six parts: five for cross-validation and one as an independent test set.

The comparative results summarized in Table 1 demonstrate that all variants of SGFF-DTA consistently achieve superior performance across both datasets. The GIN-based variant (SGFF-DTA_GIN) is particularly outstanding, surpassing all baselines across most metrics. Specifically, on the critical MSE, SGFF-DTA demonstrates superior accuracy. On Davis, it achieves a 0.173 MSE, a significant 9.4% reduction compared to the SOTA method, MambaTransDTA

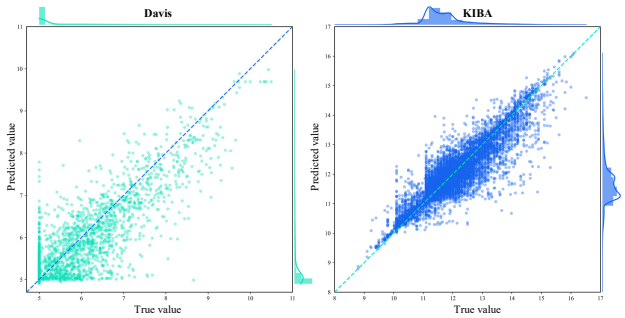

Figure 3: Affinity scatter plots on Davis and KIBA datasets.

(0.191 MSE). This improvement is primarily attributed to our deep topological exploration and efficient multimodal fusion. While MambaTransDTA captures long-range dependencies via hybrid sequence modeling, it ignores explicit topological structures. In contrast, SGFF-DTA efficiently processes protein long sequence contexts via BiMamba and extracts deep non-linear structural features using FK-AGNNs, compensating for pure sequence modeling defects.

Similarly, on the larger KIBA, the MSE of SGFF-DTA_GIN (0.125) outperforms second-best models MutualDTA (0.127) and GLCN-DTA (0.127). This advantage benefits from deeper feature interaction strategies. MutualDTA focuses on attention-based interaction, whereas SGFF-DTA constructs an interaction node within the composite graph. This allows finer sequence-graph complementation and calibration, simulating drug-target binding patterns more accurately. Furthermore, r_m^2 improvements validate this hierarchical dual-pathway architecture's robustness, reaching peak scores of 0.789 and 0.813 on Davis and KIBA, respectively. Although MutualDTA slightly leads on KIBA's CI (by 0.4%), the dual advantages of SGFF-DTA on MSE and r_m^2 demonstrate higher competitiveness in DTA prediction. To further visually validate the regression fitting capability of our model, Figure 3 plots the relationship between the test set predictions of SGFF-DTA and true values as scatter plots. For both the Davis and massive KIBA datasets, most sample points tightly converge near the ideal diagonal $y = x$, presenting significant linear correlation and extremely low dispersion. Importantly, density distribution histograms along the coordinate axes show the predicted values and true labels' distribution shapes highly overlapping. This indicates SGFF-DTA performs excellently in numerical accuracy and successfully captures global affinity distribution patterns without obvious systematic biases (e.g., overall overestimation or underestimation). This visual representation corroborates Table 1's excellent r_m^2 metrics, establishing the model's efficiency and reliability in fitting complex drug-target interactions.

Finally, we analyze SGFF-DTA framework performance under different GNN variants (GCN, GraphSAGE, GATv2, GIN). Results show that even using standard GCN operators,

the model maintains a competitive low MSE, proving the FKAGNNs framework's effectiveness independently of specific graph encoders. Among all variants, SGFF-DTA_GIN consistently performs best. Consistent with graph isomorphism theory, GIN possesses stronger structural discriminative capability than GCN and GAT (Weisfeiler-Lehman test), enabling it to acutely identify key substructural differences in the composite graph, and to provide richer topological features. Compared to GraphDTA (GIN variant) using only drug graphs, the MSE of our model on Davis and KIBA significantly decreases by 24.5% and 15.0%, respectively, fully demonstrating performance gains from introducing FKANLayer's non-linear transformations and the multimodal fusion. Given SGFF-DTA_GIN's comprehensive advantages across all metrics on benchmark datasets, we adopt this variant as the default model for subsequent experiments. Furthermore, we show that SGFF-DTA achieves improved predictive performance while maintaining competitive computational efficiency and a low memory footprint compared to baseline models (see Supplementary Material A.6 for a detailed comparison of computational costs).

3.4 GENERALIZATION ABILITY VALIDATION ON METZ AND COLD-START SETTINGS

To further verify our model's generalization across different data distributions, we expand evaluation to the Metz dataset containing diverse interactions. Following MGraphDTA settings, we randomly divide the dataset into a training set (28,207 samples) and a test set (7,052 samples), comparing it with seven advanced baselines via 5-fold cross-validation. Figure 4 shows SGFF-DTA continues its Davis and KIBA advantages on Metz, consistently outperforming comparative methods across all metrics and demonstrating strong robustness. Specifically, it achieves 0.252 MSE, 0.722 r_m^2, and 0.829 CI. Notably, compared to the second-best MGraphDTA, our model improves by 4.9%, 3.0%, and 0.85% across the three metrics, respectively, significantly surpassing the large-model-based LLMDTA. This strongly proves SGFF-DTA successfully captures universal drug-target binding rules rather than overfitting specific datasets, adapting effectively to complex biological interaction data. We report only default model results. Supplementary Table S3 presents the remaining variants' performances.

Although random splitting results are outstanding, practical applications require avoiding overly optimistic evaluations caused by exposing drug/target information to the test set [Atas Guvenilir and Doğan, 2023] . To simulate realistic drug discovery scenarios like screening novel compounds, we conduct a rigorous Cold-Start evaluation on Davis. The cold-start experiment encompasses three scenarios: Cold Drug, where test set drugs never appear in training; Cold Target, where test set proteins are completely unseen; and the most challenging Cold Pair, where both drugs and targets

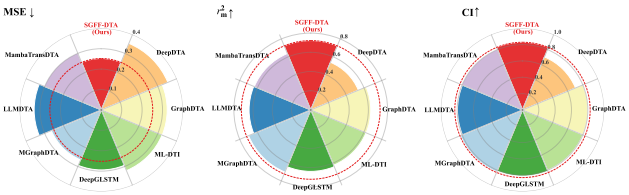

Figure 4: Comparative visualization of predictive performance on Metz. The radial bar chart illustrates SGFF-DTA and SOTA results across MSE, r_m^2, and CI. Sector radial lengths represent specific model metric values. Concentric dashed lines mark SGFF-DTA's optimal benchmark (red).

are completely new. We compare SGFF-DTA with five representative methods (ML-DTI, GraphDTA, MGraphDTA, AttentionMGT-DTA, GFLearn). All models adopt consistent 5-fold cross-validation and identical settings ensuring fairness. Supplementary Table S4 provides specific performances under these scenarios. As expected, all models' performance declines significantly compared to random splitting, corroborating the extreme difficulty of unseen domain generalization. However, our model maintains optimal comprehensive performance and robustness. Under the Cold Drug, although GFLearn slightly leads CI (by 0.3%), our model achieves a 5.9% MSE reduction, indicating SGFF-DTA predicts DTA more accurately. In the Cold Target, the model exhibits obvious superiority, reducing MSE by 12.8% and increasing r_m^2 by 6.9% compared to the second-best method. This likely results from utilizing pre-trained feature transfer to inject rich sequence semantics into graph nodes, thereby achieving a warm start that resolves unseen entity initialization difficulties. Even in the most challenging Cold Pair, SGFF-DTA maintains its lead and excellent stability. Compared to GFLearn, which also utilizes GNNs, SGFF-DTA's advantage stems from the FKANLayer. It executes complex non-linear transformations, deeply mining universal biochemical rules determining intermolecular interactions without explicit references.

Furthermore, results highlight multimodal fusion's importance for unknown data. Methods relying solely on graph features (GraphDTA, MGraphDTA) capture molecular topologies but perform poorly in cold-start scenarios lacking sequence information. Combined Metz and cold-start results fully verify SGFF-DTA's strong generalization and broad applicability across different data distributions and unknown entities.

3.5 ABLATION STUDY

To systematically verify each component's contribution, we construct variants by removing or replacing specific modules and conduct ablation experiments on Davis. Table 2 presents performance comparisons. Removing the semantic sequence representation learning module (W/o SSRL)

Table 2: Ablation Study Results of SGFF-DTA on Davis.

Model	MSE↓	r_m^2↑	CI↑
W/o SGE	0.179	0.781	0.902
W/o BiMamba	0.192	0.765	0.886
Unidirectional Mamba	0.189	0.778	0.891
W/o MHGA	0.187	0.776	0.893
W/o Interaction Node	0.184	0.769	0.897
W/o FKANLayer	0.188	0.762	0.895
W/o CGF	0.181	0.777	0.904
W/o Pre-Train	0.196	0.763	0.884
W/o SSRL	0.227	0.694	0.872
W/o CGRL	0.204	0.743	0.876
SGFF-DTA	**0.173**	**0.789**	**0.910**

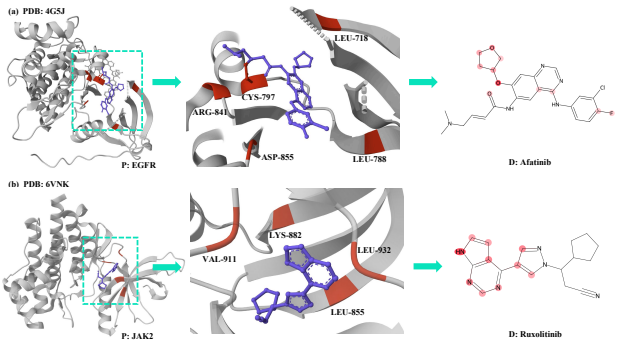

Figure 5: Visualization and interpretability analysis of attention weights in drug-target complexes. Left: 3D complex structures (drug: purple; protein: gray). Middle: Zoomed-in 3D binding pockets. Right: 2D drug molecular graphs.

severely damages performance, which causes MSE to surge to 0.227 (a performance drop of 31.2%). In this variant, we use only the graph modality, adopting human-defined features for node initialization instead of sequence-encoder-extracted features. This huge gap proves SSRL's contextual semantics contain rich biochemical information unmatched by handcrafted features. Similarly, removing the composite graph representation learning module (W/o CGRL), retaining only the sequence modality, causes a 17.9% performance drop. This indicates relying solely on sequences fails to capture molecular topologies and non-linear interactions. Therefore, deep sequence-graph synergy is a prerequisite for high-precision DTA prediction. Regarding the sequence encoder, replacing BiMamba with BiLSTM (W/o BiMamba) causes an 11.0% degradation. This gap stems from BiMamba's superiority capturing global protein dependencies, overcoming RNN gradient bottlenecks. To further isolate the contribution of bidirectionality, we evaluate a variant employing standard Unidirectional Mamba, which increases the MSE to 0.189. This rigorous comparison demonstrates that the spatial dependencies of protein sequences are non-causal, and standard causal scanning fails to capture essential global context from the C-terminus. For the graph encoder FKAGIN, reverting to the original GIN (W/o FKANLayer) degrades MSE by 8.7%. This confirms complex non-linear Fourier basis transformations acutely mine implicit high-order topological patterns in the composite graph. Furthermore, removing MHGA, interaction node, and SGE modules degrades performance, verifying the necessity of explicit cross-modal bridging, feature denoising, and enhancement.

The W/o Pre-Train variant demonstrates pre-trained feature transfer strategy importance. Changing to end-to-end training, directly transferring untrained sequence encoder features to graph nodes as initial states, decreases model performance by 13.3%. This shows pre-training is not mere parameter initialization but a crucial semantic alignment process. It constructs a composite graph feature space with rich biochemical priors, preventing GNNs from falling into early suboptimal solutions lacking node semantics, enabling focus on mining complex topological patterns. Finally, replacing the CGF module with simple concatenation causes a 4.6% drop, indicating CGF's mutual perception gating dynamically screens and integrates multimodal information, promoting inter-modal fusion. Ablation results show the complete model achieves optimal performance across all metrics. Removing any component causes accuracy loss (3.5%-31.2%), reflecting functional complementarity and fully verifying the overall architectural design.

4 INTERPRETABILITY ANALYSIS AND VISUALIZATION

In DTA prediction, interpretability verifies whether models learn biologically meaningful features rather than exploiting data biases. We select two representative kinase-inhibitor PDB co-crystal structures for visualization: EGFR-Afatinib (4G5J) [Solca et al., 2012] and JAK2-Ruxolitinib (6VNK) [Davis et al., 2021]. Mapping MHGA attention weights onto 3D protein structures and 2D drug molecular graphs, red-highlighted regions represent high-attention residues or atoms. We compare the top 15 predicted high-weight residues with experimentally observed binding residues within a 5 Å ligand-centered radius. In EGFR-Afatinib (Figure 5 a), the model accurately captures binding pocket residues (CYS-797, LEU-718, ARG-841, ASP-855, LEU-788), confirmed by crystallographic studies to bind via hydrogen bonds or hydrophobic interactions. Notably, CYS-797 is crucial for Afatinib covalent bonding. Similarly, for JAK2-Ruxolitinib (Figure 5 b), highly attended residues (LYS-882, LEU-932, VAL-911, LEU-855) overlap with the crystal structure-defined ATP binding pocket. This spatial consistency shows MHGA autonomously extracts interaction information without explicit 3D coordinate supervision.

Furthermore, visualizing 2D drug molecular graphs, darker atoms represent higher attention weights. The model concentrates attention on N/O-rich heterocyclic structures and key functional groups. For Afatinib, the model focuses on the central quinazoline ring and side-chain O atoms, typical backbones for intermolecular hydrogen bonds and $\pi-\pi$ stacking. For Ruxolitinib, attention concentrates on the pyrrolo[2,3-d]pyrimidine backbone and connecting nitrogen atoms, aligning with JAK2 key inhibitor hinge regions. In summary, our model accurately identifies protein pocket binding sites and key affinity-contributing functional groups, providing an interpretable method for drug discovery. A rigorous quantitative evaluation of the attention mechanism (Precision@K) is provided in Supplementary Material A.7.

5 CONCLUSION

This study proposes SGFF-DTA, a sequence-graph fusion framework addressing sequence long-range dependencies and the fitting of complex non-linear topologies. Through a mathematically complementary synergy, it utilizes Bi-Mamba to extract non-causal global sequence context at linear complexity, while employing Fourier-KAN to project graph features into the spectral space, executing learnable non-linear transformations that effectively break GNNs expressivity limits. Relying on interaction node and cross-gated fusion, the framework achieves deep drug-protein cross-modal synergy. Furthermore, pre-trained node feature transfer significantly improves robustness in information-scarce scenarios. Comprehensive experiments on Davis, KIBA, and Metz demonstrate that SGFF-DTA achieves SOTA predictive performance and exceptional generalization. Ablation studies highlight component contributions, while case visualization confirms interpretability in locating key binding sites and pharmacophores.

Despite the excellent performance, certain limitations remain. Relying on static residue contact maps and sequence features limits the sensitivity to complex dynamic conformational changes during binding. Furthermore, zero-shot prediction for novel orphan targets may be constrained by data scarcity. Future work will focus on introducing geometric deep learning for spatiotemporal binding dynamics [Min et al., 2024, Nguyen and Kang, 2025] and exploring the knowledge enhancement strategies of large language models (LLM) [Soman et al., 2024].

Acknowledgements

This work was supported by Basic Research Program of Shanxi Province (202303021211069 and 202403021222070), and Key Research and Development Program of Shanxi Province (202402020101008). We would like to thank the anonymous reviewers and the meta-reviewer for their valuable comments and suggestions.

References

Karim Abbasi, Parvin Razzaghi, Antti Poso, Massoud Amanlou, Jahan B Ghasemi, and Ali Masoudi-Nejad. DeepCDA: Deep cross-domain compound–protein affinity prediction through LSTM and convolutional neural networks. *Bioinformatics*, 36(17):4633–4642, 2020.

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024.

Shadab Ahmad, Leonardo Jose da Costa Gonzales, Emily H Bowler-Barnett, Daniel L Rice, Minjoon Kim, Supun Wijerathne, Aurélien Luciani, Swaathi Kandasaamy, Jie Luo, Xavier Watkins, et al. The UniProt website API: Facilitating programmatic access to protein knowledge. *Nucleic Acids Research*, 53(W1):W547–W553, 2025.

Heval Atas Guvenilir and Tunca Doğan. How to approach machine learning-based prediction of drug/compound–target interactions. *Journal of Cheminformatics*, 15(1): 16, 2023.

Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? In *International Conference on Learning Representations*, 2022.

Mindy I Davis, Jeremy P Hunt, Sanna Herrgard, Pietro Ciceri, Lisa M Wodicka, Gabriel Pallares, Michael Hocker, Daniel K Treiber, and Patrick P Zarrinkar. Comprehensive analysis of kinase inhibitor selectivity. *Nature Biotechnology*, 29(11):1046–1051, 2011.

Ryan R Davis, Baoli Li, Sang Y Yun, Alice Chan, Pradeep Nareddy, Steven Gunawan, Muhammad Ayaz, Harshani R Lawrence, Gary W Reuther, Nicholas J Lawrence, et al. Structural insights into JAK2 inhibition by ruxolitinib, fedratinib, and derivatives thereof. *Journal of Medicinal Chemistry*, 64(4):2228–2241, 2021.

Jiejin Deng, Yijia Zhang, Yaohua Pan, Xiaobo Li, and Mingyu Lu. MultiDTA: Drug-target binding affinity prediction via representation learning and graph convolutional neural networks. *International Journal of Machine Learning and Cybernetics*, 15(7):2709–2718, 2024.

Piero Fariselli, Osvaldo Olmea, Alfonso Valencia, and Rita Casadio. Prediction of contact maps with neural networks and correlated mutations. *Protein Engineering*, 14(11): 835–843, 2001.

Jennifer Fleming, Paulyna Magana, Sreenath Nair, Maxim Tsenkov, Damian Bertoni, Ivanna Pidruchna, Marcelo Querino Lima Afonso, Adam Midlik, Urmila Paramval, Augustin Žídek, et al. AlphaFold protein structure database and 3D-beacons: New data and capabilities. *Journal of Molecular Biology*, 437(15):168967, 2025.

Mithat Gönen and Glenn Heller. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, 92(4):965–970, 2005.

Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling (COLM)*, 2024.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30, 2017.

Haohuai He, Guanxing Chen, Zhenchao Tang, and Calvin Yu-Chian Chen. Dual modality feature fused neural network integrating binding site information for drug target affinity prediction. *NPJ Digital Medicine*, 8(1):67, 2025.

Tong He, Marten Heidemeyer, Fuqiang Ban, Artem Cherkasov, and Martin Ester. SimBoost: A read-across approach for predicting drug–target binding affinities using gradient boosting machines. *Journal of Cheminformatics*, 9(1):24, 2017.

Zibo Huang, Xinrui Weng, and Le Ou-Yang. GFLearn: Generalized feature learning for drug-target binding affinity prediction. *IEEE Journal of Biomedical and Health Informatics*, 2025.

Yanlei Kang, Haoyu Zhuang, Yunliang Jiang, and Zhong Li. MF-DTA: Predicting drug-target affinity with multimodal feature fusion model. *Journal of Biomedical Informatics*, page 104926, 2025.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Radoslav Krivák and David Hoksza. P2Rank: Machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics*, 10(1):39, 2018.

Greg Landrum et al. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum*, 8(31.10):5281, 2013.

Wenjun Li, Yiqiang Zhou, and Xiwei Tang. TF-DTA: A deep learning approach using transformer encoder to predict drug-target binding affinity. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2023.

Yuxuan Li, Xiang Li, and Jian Yang. Spatial group-wise enhance: Enhancing semantic feature learning in cnn. In *Proceedings of the Asian Conference on Computer Vision*, 2022.

Xuan Lin, Kaiqi Zhao, Tong Xiao, Zhe Quan, Zhi-Jie Wang, and Philip S Yu. DeepGS: Deep representation learning of graphs and sequences for drug-target binding affinity prediction. In *24th European Conference on Artificial Intelligence (ECAI)*, 2020.

Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljacic, Thomas Hou, and Max Tegmark. KAN: Kolmogorov–Arnold Networks. In *International Conference on Learning Representations*, 2025.

Xinpo Lou, Jianxiu Cai, Qidong Liu, and Shirley WI Siu. MambaTransDTA: A hybrid Mamba-Transformer architecture for accurate drug-target binding affinity prediction. *Journal of Chemical Information and Modeling*, 66(1): 259–270, 2025.

Alfonso Eduardo Marquez-Chamorro, Gualberto Asencio-Cortes, Federico Divina, and Jesus Salvador Aguilar-Ruiz. Evolutionary decision rules for predicting protein contact maps. *Pattern Analysis and Applications*, 17(4):725–737, 2014.

James T Metz, Eric F Johnson, Niru B Soni, Philip J Merta, Lemma Kifle, and Philip J Hajduk. Navigating the kinome. *Nature Chemical Biology*, 7(4):200–202, 2011.

Yaosen Min, Ye Wei, Peizhuo Wang, Xiaoting Wang, Han Li, Nian Wu, Stefan Bauer, Shuxin Zheng, Yu Shi, Yingheng Wang, et al. From static to dynamic structures: Improving binding affinity prediction with graph-based deep learning. *Advanced Science*, 11(40):2405404, 2024.

Shrimon Mukherjee, Madhusudan Ghosh, and Partha Basuchowdhuri. DeepGLSTM: Deep graph convolutional network and LSTM based approach for predicting drug-target binding affinity. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, 2022.

Ngoc-Quang Nguyen and Jaewoo Kang. EquiCPI: SE (3)-equivariant geometric deep learning for structure-aware prediction of compound-protein interactions. *Journal of Chemical Information and Modeling*, 65(13):7252–7262, 2025.

Thin Nguyen, Hang Le, Thomas P Quinn, Tri Nguyen, Thuc Duy Le, and Svetha Venkatesh. GraphDTA: Predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, 37(8):1140–1147, 2021.

Hakime Öztürk, Arzucan Özgür, and Elif Ozkirimli. DeepDTA: Deep drug–target binding affinity prediction. *Bioinformatics*, 34(17):i821–i829, 2018.

Tapio Pahikkala, Antti Airola, Sami Pietilä, Sushil Shakyawar, Agnieszka Szwajda, Jing Tang, and Tero Aittokallio. Toward more realistic drug–target interaction predictions. *Briefings in Bioinformatics*, 16(2):325–337, 2015.

Partha Pratim Roy, Somnath Paul, Indrani Mitra, and Kunal Roy. On two novel parameters for validation of predictive QSAR models. *Molecules*, 14(5):1660–1701, 2009.

Haiou Qi, Ting Yu, Wenwen Yu, and Chenxi Liu. Drug–target affinity prediction with extended graph learning-convolutional networks. *BMC Bioinformatics*, 25(1):75, 2024.

Chayna Sarkar, Biswadeep Das, Vikram Singh Rawat, Julie Birdie Wahlang, Arvind Nongpiur, Iadarilang Tiewsoh, Nari M Lyngdoh, Debasmita Das, Manjunath Bidarolli, and Hannah Theresa Sony. Artificial intelligence and machine learning technology driven modern drug discovery and development. *International Journal of Molecular Sciences*, 24(3):2026, 2023.

Flavio Solca, Goeran Dahl, Andreas Zoephel, Gerd Bader, Michael Sanderson, Christian Klein, Oliver Kraemer, Frank Himmelsbach, Eric Haaksma, and Guenther R Adolf. Target binding properties and cellular activity of afatinib (BIBW 2992), an irreversible ErbB family blocker. *The Journal of Pharmacology and Experimental therapeutics*, 343(2):342–350, 2012.

Karthik Soman, Peter W Rose, John H Morris, Rabia E Akbas, Brett Smith, Braian Peetoom, Catalina Villouta-Reyes, Gabriel Cerono, Yongmei Shi, Angela Rizk-Jackson, et al. Biomedical knowledge graph-optimized prompt generation for large language models. *Bioinformatics*, 40(9):btae560, 2024.

Jing Tang, Agnieszka Szwajda, Sushil Shakyawar, Tao Xu, Petteri Hintsanen, Krister Wennerberg, and Tero Aittokallio. Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis. *Journal of Chemical Information and Modeling*, 54 (3):735–743, 2014.

Wuguo Tang, Qichang Zhao, and Jianxin Wang. LLMDTA: Improving cold-start prediction in drug-target affinity with biological LLM. *IEEE Transactions on Computational Biology and Bioinformatics*, 2025.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

Roger Waleffe, Wonmin Byeon, Duncan Riach, Brandon Norick, Vijay Korthikanti, Tri Dao, Albert Gu, Ali Hatamizadeh, Sudhakar Singh, Deepak Narayanan, et al. An empirical study of Mamba-based language models. *arXiv preprint arXiv:2406.07887*, 2024.

Xun Wang, Zhijun Xia, Runqiu Feng, Tongyu Han, Hanyu Wang, Wenqian Yu, and Xingguang Wang. SMFF-DTA: Using a sequential multi-feature fusion method with multiple attention mechanisms to predict drug-target binding affinity. *BMC Biology*, 23(1):120, 2025.

Olivier J Wouters, Martin McKee, and Jeroen Luyten. Research and development costs of new drugs—reply. *The Journal of the American Medical Association*, 324(5): 518–518, 2020.

Hongjie Wu, Junkai Liu, Tengsheng Jiang, Quan Zou, Shujie Qi, Zhiming Cui, Prayag Tiwari, and Yijie Ding. AttentionMGT-DTA: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism. *Neural Networks*, 169:623–636, 2024a.

Li Wu, Wenbin Pei, Jiulong Jiao, and Qiang Zhang. UmambaTSF: A U-shaped multi-scale long-term time series forecasting method using Mamba. *arXiv preprint arXiv:2410.11278*, 2024b.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

Ziduo Yang, Weihe Zhong, Lu Zhao, and Calvin Yu-Chian Chen. ML-DTI: Mutual learning mechanism for interpretable drug–target interaction prediction. *The Journal of Physical Chemistry Letters*, 12(17):4247–4261, 2021.

Ziduo Yang, Weihe Zhong, Lu Zhao, and Calvin Yu-Chian Chen. MGraphDTA: Deep multiscale graph neural network for explainable drug–target binding affinity prediction. *Chemical Science*, 13(3):816–833, 2022.

Yongna Yuan, Siming Chen, Rizhen Hu, and Xin Wang. MutualDTA: An interpretable drug–target affinity prediction model leveraging pretrained models and mutual attention. *Journal of Chemical Information and Modeling*, 65(3): 1211–1227, 2025.

Xin Zeng, Shu-Juan Li, Shuang-Qing Lv, Meng-Liang Wen, and Yi Li. A comprehensive review of the recent advances on predicting drug-target affinity based on deep learning. *Frontiers in Pharmacology*, 15:1375522, 2024.

Xingran Zhao, Yanbu Guo, Bingyi Wang, and Weihua Li. CKDTA: A chemical knowledge-enhanced framework for drug-target affinity prediction. *Journal of Computational Science*, page 102706, 2025.

A Sequence-Graph Fusion Framework via BiMamba and Fourier-KAN for Interpretable Drug-Target Affinity Prediction (Supplementary Material)

Xibo Li[1] **Lian Chen**[1] **Dingyuan Chen**[1] **Yichuan Zhao**[1] **Li Zhou**[*2] **Dongxi Li**[*2]

[1]College of Artificial Intelligence, Taiyuan University of Technology, Shanxi, China
[2]College of Computer Science and Technology, Taiyuan University of Technology, Shanxi, China
[*]Corresponding authors: zhouli@tyut.edu.cn, dxli0426@126.com

A SUPPLEMENTARY TABLES

A.1 DETAILED DATASETS STATISTICS

We evaluate the predictive performance of the proposed SGFF-DTA model using three widely recognized benchmark datasets: Davis [Davis et al., 2011], KIBA [Tang et al., 2014], and Metz [Metz et al., 2011]. Table S1 provides a comprehensive statistical breakdown of these datasets.

Specifically, the table details the number of unique drugs and targets, the total number of experimentally validated interactions, the graph density of the interaction network, and the distribution range of the affinity values. Notably, the Davis dataset presents a fully dense interaction matrix (100% density), whereas KIBA and Metz exhibit significantly lower density (24.46% and 14.58%, respectively). This varied distribution of graph density comprehensively tests the capacity of the model to generalize across both highly dense and sparse interaction topologies.

Table S1: Detailed Statistics of the Benchmark Datasets.

Dataset	Drugs	Targets	Interactions	Density	Value Range
Davis	68	442	30,056	100%	[5.0, 10.8]
KIBA	2,111	229	118,254	24.46%	[0.0, 17.2]
Metz	1,423	170	35,259	14.58%	[0.0, 12.0]

A.2 DETAILED HYPERPARAMETER CONFIGURATIONS

The optimal hyperparameters for SGFF-DTA were determined through a comprehensive grid search and 5-fold cross-validation exclusively on the Davis dataset. To rigorously evaluate the generalization capabilities of the model, these identical settings were subsequently applied to the KIBA and Metz datasets.

Table S2 comprehensively lists the default values for all key hyperparameters utilized during the training phase. The table details the optimization strategy, including the specific initial and minimum learning rates controlled by the CosineAnnealingLR scheduler. Furthermore, it explicitly outlines the structural dimensions of our core modules, such as the state and convolution dimensions for the BiMamba sequence encoder, and the grid size and layer depth for the FKAGIN graph representation module.

A.3 PERFORMANCE OF SGFF-DTA VARIANTS ON THE METZ DATASET

We further evaluated our model on the Metz dataset to validate its generalization ability. While the main manuscript reports the results of our default model, Table S3 details the performance of all four SGFF-DTA variants (GCN, GraphSAGE,

Table S2: Detailed Hyperparameter Configurations for the SGFF-DTA Model.

Hyperparameter	Default value
Optimizer	Adam
Schedule	CosineAnnealingLR
LR_{max}	$1e^{-3}$
LR_{min}	$1e^{-6}$
T_0	36
T_{mult}	2
Pre-training epochs	150
Training epochs	200
Batch size	256
Dropout	0.2
Attention heads	8
Early stopping	40
SGE_G	16
Embedding dimension	256
Mamba d_{model}	128
Mamba d_{state}	32
Mamba d_{conv}	4
Mamba expand	2
FourierKAN grid_size	5
FKAGIN layers	4
FKAGIN dimension	128

GATv2, and GIN).

Consistent with the observations on the Davis and KIBA datasets, the SGFF-DTA_GIN achieves the most optimal predictive performance across all three evaluation metrics. Nevertheless, all variants demonstrate highly competitive and stable results, confirming that the robustness of our architecture stems from the core sequence-graph fusion mechanism rather than reliance on a specific graph convolution operator.

Table S3: Performance Comparison of SGFF-DTA Variants on the Metz Dataset.

Method	MSE↓	r_m^2↑	CI↑
SGFF-DTA_GCN	0.255	<u>0.719</u>	<u>0.827</u>
SGFF-DTA_GraphSAGE	<u>0.253</u>	0.716	0.824
SGFF-DTA_GATv2	0.259	0.711	0.822
SGFF-DTA_GIN	**0.252**	**0.722**	**0.829**

Note: Bold values indicate the best performance, and underlined values indicate the second best. ↓ denotes lower is better; ↑ denotes higher is better.

A.4 DETAILED RESULT OF COLD-START EVALUATION

As discussed in the main manuscript, practical drug discovery frequently involves screening novel compounds or unseen protein targets. To simulate these realistic out-of-distribution scenarios, we conducted a rigorous cold-start evaluation on the Davis dataset.

Table S4 presents the detailed quantitative results across three distinct challenging settings: *Cold Drug* (test drugs are entirely unseen during training), *Cold Target* (test proteins are completely unseen), and the most extreme *Cold Pair* (both drugs and targets are new). While all baseline models exhibit significant performance degradation compared to the random splitting task, the proposed SGFF-DTA consistently maintains optimal predictive accuracy and robustness across almost all metrics, demonstrating its superior generalization capability for unknown biochemical entities.

Table S4: Performance Comparison Under Three Cold-Start Scenarios on the Davis Dataset.

Scenario	Method	MSE↓	r_m^2↑	CI↑
Cold Drug	ML-DTI	0.712	0.179	0.660
	GraphDTA	0.695	0.198	0.692
	MGraphDTA	0.683	0.242	0.704
	AttentionMGT-DTA	0.654	0.213	0.717
	GFLearn (2025)	<u>0.572</u>	<u>0.251</u>	**0.728**
	SGFF-DTA (ours)	**0.538**	**0.284**	<u>0.726</u>
Cold Target	ML-DTI	0.497	0.244	0.716
	GraphDTA	0.513	0.231	0.724
	MGraphDTA	0.461	0.285	0.767
	AttentionMGT-DTA	0.396	0.437	<u>0.832</u>
	GFLearn (2025)	<u>0.375</u>	<u>0.461</u>	0.826
	SGFF-DTA (ours)	**0.327**	**0.493**	**0.849**
Cold Pair	ML-DTI	0.759	0.074	0.583
	GraphDTA	0.783	0.067	0.609
	MGraphDTA	0.743	0.079	0.616
	AttentionMGT-DTA	<u>0.641</u>	0.095	0.625
	GFLearn (2025)	0.657	<u>0.104</u>	<u>0.643</u>
	SGFF-DTA (ours)	**0.615**	**0.118**	**0.661**

Note: Bold values indicate the best performance, and underlined values indicate the second best. ↓ denotes lower is better; ↑ denotes higher is better.

A.5 ABLATION STUDY OF FKANLAYER ACROSS DIFFERENT VARIANTS

In the main manuscript, we reported the ablation study results primarily based on our default architecture (utilizing GIN). To demonstrate that the performance gain achieved by the Fourier-KAN layer is universal and not restricted to a specific graph convolution operator, we extended the ablation experiments across all four variants (GCN, GraphSAGE, GATv2, and GIN) on the Davis dataset.

Table S5 details these comprehensive comparative results. For each baseline variant, replacing the FKANLayer with traditional linear projections (denoted as W/o FKANLayer) consistently results in a noticeable degradation across all three metrics. This universal improvement firmly confirms that the non-linear spectral transformations introduced by the Fourier basis functions can robustly capture complex high-order topological patterns, effectively benefiting diverse graph representation learning architectures.

Table S5: Ablation Study of FKANLayer Across Different Variants on the Davis Dataset.

Model	MSE↓	r_m^2↑	CI↑
W/o FKANLayer	0.192	0.763	0.892
SGFF-DTA_GCN	**0.176**	**0.784**	**0.902**
W/o FKANLayer	0.197	0.753	0.894
SGFF-DTA_GraphSAGE	**0.178**	**0.778**	**0.905**
W/o FKANLayer	0.189	0.758	0.895
SGFF-DTA_GATv2	**0.175**	**0.782**	**0.908**
W/o FKANLayer	0.188	0.762	0.895
SGFF-DTA_GIN	**0.173**	**0.789**	**0.910**

Note: Bold values indicate the performance of the complete model equipped with the FKANLayer for each respective GNN variant. ↓ denotes lower is better; ↑ denotes higher is better.

A.6 COMPUTATIONAL EFFICIENCY AND RESOURCE CONSUMPTION COMPARISON

To evaluate the computational efficiency and resource overhead of the proposed SGFF-DTA model when handling long sequences and large-scale graph structures, we conduct a quantitative comparison against five representative SOTA baselines (MultiDTA, MGraphDTA, AttentionMGT-DTA, MutualDTA, and MambaTransDTA) on the Davis dataset. The models are evaluated across three key metrics: training runtime per epoch (seconds), peak GPU memory footprint (MB), and Mean Squared Error (MSE) on the test set. To ensure a fair comparison, all experiments are conducted under identical hardware and software environments.

As shown in Table S6, SGFF-DTA demonstrates clear advantages in both computational efficiency and resource consumption. Specifically, traditional Transformer-based models, such as AttentionMGT-DTA, MutualDTA, and MambaTransDTA, exhibit a quadratic time complexity of $O(L^2)$ because the self-attention mechanism scales quadratically when processing long protein sequences. Consequently, these models require long training times (up to 274.93 seconds per epoch) and incur high GPU memory overhead (exceeding 21 GB). In contrast, SGFF-DTA utilizes a BiMamba encoder with a linear time complexity of $O(L)$ to model the global context of protein sequences through bidirectional selective state-space scanning. This design reduces the training runtime to 35.80 seconds per epoch, which represents an approximate eight-fold acceleration, while maintaining a lower GPU memory footprint of 14,239 MB. This result demonstrates the feasibility of overcoming the computational complexity bottleneck of the Transformer in practical applications. Furthermore, compared to the lightweight, graph-only model MGraphDTA (27.26 seconds per epoch and 4,325 MB of GPU memory), SGFF-DTA introduces a minor runtime overhead of only 8.54 seconds per epoch, while achieving a 16.4% relative reduction in the test MSE (from 0.207 to 0.173). This trade-off between accuracy and efficiency validates the synergistic effect of integrating the Fourier-KAN for non-linear topological transformations and applying cross-gated fusion (CGF). Overall, SGFF-DTA achieves competitive predictive performance while maintaining low computational and memory overhead, demonstrating its capability to handle complex molecular topologies and sequence dependencies.

Table S6: Comparison of Computational Efficiency and Resource Consumption on the Davis Dataset.

Method	Runtime↓	GPU Consumption↓	MSE↓
MultiDTA	46.74	<u>7,462</u>	0.231
MGraphDTA	**27.26**	**4,325**	0.207
AttentionMGT-DTA	274.93	22,683	0.193
MutualDTA	139.07	21,691	0.196
MambaTransDTA	193.15	23,548	<u>0.191</u>
SGFF-DTA (ours)	<u>35.80</u>	14,239	**0.173**

Note: Bold values indicate the best performance, and underlined values indicate the second best. ↓ denotes lower is better.

A.7 QUANTITATIVE EVALUATION OF MODEL INTERPRETABILITY

Although qualitative visualizations of individual cases (such as EGFR-Afatinib and JAK2-Ruxolitinib) demonstrate that the multi-head gated attention (MHGA) module focuses on known active sites, we also perform a systematic, macro-level quantitative evaluation to establish a rigorous statistical link between the attention mechanism of the model and experimental structural biology.

Specifically, we randomly select ten drug-target complexes from the independent test set that have experimentally resolved 3D co-crystal structures available in the Protein Data Bank (PDB). For each complex, we define the ground-truth binding pocket residues as any protein residue located within a 5 Å spatial distance from any atom of the bound drug molecule. To evaluate the alignment of the attention weights of the model with these experimental binding sites, we extract the sequence-level attention weights generated by the MHGA module for the target protein. We rank all residues according to their attention weights and select the Top-K residues (where $K \in \{10, 20\}$) with the highest weights as the predicted binding residues. We then compute the Precision@K metric, which measures the proportion of predicted Top-K residues that fall within the ground-truth binding pocket.

Table S7 compares the Precision@10 and Precision@20 of our model against the random guess baseline. SGFF-DTA achieves a Precision@10 of 35.6% and a Precision@20 of 33.8%, demonstrating that more than a third of the highly-attended residues directly correspond to the physical binding pocket. Given that the average length of the target proteins exceeds 600

residues, the physical binding pocket represents only a small fraction of the sequence, resulting in an expected Precision@K of less than 5.0% for the random guess baseline. The precision of the proposed model exceeds this baseline by more than seven-fold, which statistically validates that SGFF-DTA autonomously and accurately concentrates its attention on experimentally verified binding pockets. Notably, this targeted alignment is learned entirely from sequence and graph topology under weak supervision, without any explicit 3D spatial coordinate supervision during training. This quantitative result suggests that SGFF-DTA captures genuine physicochemical rules of drug-target affinity rather than exploiting dataset shortcut biases.

Table S7: Quantitative Interpretability Performance of SGFF-DTA on Selected PDB Complexes.

Method	Precision@10 (%)	Precision@20 (%)
Random Guess	< 5.0	< 5.0
SGFF-DTA (ours)	**35.6**	**33.8**

A.8 PARAMETER-MATCHED ABLATION STUDY AND SPECTRAL COEFFICIENT ANALYSIS

To clarify whether the performance improvement of the FKAGIN module originates from the spectral space transformation of the Fourier basis or from an increase in the capacity of the model, we conducted a systematic parameter-matched ablation study on the Davis dataset.

In our architecture, the FKANLayer performs learnable non-linear transformations by expanding features via multiple Fourier basis functions (harmonics), which typically requires more parameters than a simple linear transformation. Consequently, replacing the FKANLayer with a standard MLP reduces the total parameter count. To ensure a fair evaluation, we designed a Wider-MLP baseline by widening the hidden dimensions of the standard MLP, thereby scaling its parameter count to match or slightly exceed that of the proposed FKAGIN model. Additionally, we implemented a KAGIN baseline based on the original B-spline KAN, ensuring its parameters are aligned with those of our method.

Table S8 presents the results of the parameter-matched comparison. The results show that increasing the parameter capacity of the MLP network (Wider-MLP) brings performance improvements compared to the baseline MLP. However, its performance quickly encounters a bottleneck and remains lower than that of the proposed model. In contrast, under the condition of matched parameter capacity, the proposed FKAGIN model achieves a lower MSE. Furthermore, Fourier-KAN outperforms the traditional B-spline KAN. This controlled evaluation demonstrates that the performance of the proposed model stems from the inductive bias of the Fourier basis rather than an increase in the number of parameters.

Furthermore, although traditional message-passing layers act as low-pass filters that tend to cause feature over-smoothing in deep GNNs, the FKANLayer functions as an adaptive spectral filter. By projecting features into the spectral domain, the Fourier basis functions naturally capture both low-frequency global graph topologies and high-frequency local structural variations, thereby alleviating the bottleneck of over-smoothing.

Table S8: Performance Comparison of Parameter-Matched Baselines on the Davis Dataset.

Model	MSE↓	r_m^2↑	CI↑
Baseline MLP	0.188	0.769	0.894
Wider-MLP	<u>0.182</u>	0.774	<u>0.901</u>
KAGIN (B-Spline-KAN)	0.183	<u>0.776</u>	0.897
FKAGIN (Ours)	**0.173**	**0.789**	**0.910**

Note: Bold values indicate the best performance, and underlined values indicate the second best. ↓ denotes lower is better; ↑ denotes higher is better.

Qualitative Analysis of Learned Spectral Coefficients. To verify the mechanism of the Fourier basis, we analyze the learned spectral coefficients $(w_{c,k}, w_{s,k})$ in the trained FKAGIN network. Setting the Fourier truncation frequency to $G = 5$, we observe the average amplitude distribution of the activated coefficients on the graph nodes. The dominant low-frequency coefficients (such as $k = 1$ and $k = 2$) smoothly capture global graph homophily and backbone topology. Notably, the high-frequency coefficients (such as $k = 4$ and $k = 5$) do not decay to zero; they remain active to capture

localized, non-linear topological variations (such as non-isomorphic pharmacophores) that are often prone to over-smoothing in standard MLPs. This distinct frequency distribution illustrates how Fourier-KAN dynamically adapts to the complex biochemical interactions.

B SUPPLEMENTARY NOTES

B.1 IMPLEMENTATION DETAILS OF FKAGNNS

As discussed in the main text, the proposed FKANLayer $\phi(x)$ is highly versatile and architecture-agnostic. While we primarily focus on the FKAGIN model to achieve the best experimental performance, this learnable non-linear transformation can be naturally integrated into other classic graph neural network architectures, forming a general framework named FKAGNNs. By replacing traditional linear projections with our proposed FKANLayer, we can fundamentally enhance the expressiveness and function-fitting capabilities of these baseline models. Below, we detail the specific implementations of three extended architectures: Graph Attention Networks v2 (FKAGATv2), Graph Convolutional Networks (FKAGCN), and GraphSAGE (FKAGraphSAGE).

FKAGATv2. Graph Attention Networks v2 (GATv2) [Brody et al., 2022] improves upon the standard GAT [Veličković et al., 2018] by introducing a dynamic attention mechanism. While the standard GAT computes attention scores using a static ranking that is independent of the query node, GATv2 evaluates the attention vector after combining the interacting node features. For Drug-Target Affinity (DTA) prediction, this dynamic property is highly beneficial. The interaction strength between a specific drug atom and a protein residue is highly context-dependent; GATv2 allows the model to adaptively evaluate the importance of local topological interactions based on the unique states of specific node pairs.

To further strengthen this dynamic learning process within the composite graph, we propose FKAGATv2. We replace the traditional linear feature projections with independent FKANLayers for the target node (ϕ_l) and the source node (ϕ_r). This mechanism maps the interacting features into a highly expressive spectral space before calculating their correlation. The attention score e_{ij} between node i and its neighbor j is computed as:

$$e_{ij} = \mathbf{a}^\top \left(\phi_l(h_i^{(l-1)}) + \phi_r(h_j^{(l-1)}) \right), \tag{18}$$

where $\mathbf{a}$ is a learnable attention vector. These scores are then normalized across all neighbors $j \in \mathcal{N}(i)$ using the softmax function to obtain the dynamic attention weights α_{ij}.

Following the weight calculation, the neighborhood features are aggregated. Uniquely in our design, the aggregated representation undergoes a secondary non-linear refinement. After the multi-head attention aggregation, the combined output is fed into an additional FKANLayer (ϕ_{update}). Assuming K independent attention heads, the final node update at the l-th layer is formulated as:

$$h_i^{(l)} = \phi_{update} \left(\Big\Vert_{k=1}^{K} \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^k \phi_r^k \left(h_j^{(l-1)} \right) \right), \tag{19}$$

where $[\ \Vert\ ]$ denotes the concatenation operation. This dual-stage spectral mapping mechanism enables FKAGATv2 to deeply and flexibly capture the complex, dynamic spatial dependencies between drugs and targets, significantly boosting the predictive capacity of the model.

FKAGCN. Graph Convolutional Networks (GCN) [Kipf and Welling, 2017] capture topological structures by isotropically aggregating neighborhood features through a symmetric normalized adjacency matrix. In a standard GCN, the message passing and feature transformation heavily rely on a learnable linear weight matrix W. However, this linear projection paradigm restricts the capacity of the model to represent highly complex structural patterns within the composite graph.

To overcome this expressiveness bottleneck, we propose FKAGCN by integrating the FKANLayer into the GCN architecture. Specifically, we replace the conventional linear feature transformation with the Fourier-KAN mapping function ϕ, empowering the model to execute spectral-level non-linear transformations on the node features. By adding self-loops to retain the intrinsic properties of the central node, the node update rule at the l-th layer of FKAGCN is formulated as follows:

$$h_v^{(l)} = \sum_{u \in \mathcal{N}(v) \cup \{v\}} \frac{1}{\sqrt{\tilde{d}_v \tilde{d}_u}} \phi \left(h_u^{(l-1)} \right), \tag{20}$$

where $\mathcal{N}(v) \cup \{v\}$ denotes the local neighborhood of node v including the node itself, while $\tilde{d}_v$ and $\tilde{d}_u$ represent the structural degrees of nodes v and u calculated from the adjacency matrix with added self-loops. By incorporating the FKANLayer into the symmetric normalized aggregation process, FKAGCN effectively maps the smoothed neighborhood features into a highly expressive latent space.

FKAGraphSAGE. GraphSAGE [Hamilton et al., 2017] is a highly effective inductive learning framework. Instead of training embeddings for each node directly, it learns a function that generates embeddings by sampling and aggregating features from the local neighborhood of a node. In the standard GraphSAGE architecture, the aggregated neighbor features are concatenated with the current features of the central node, followed by a standard linear weight transformation.

To better capture the highly complex spatial dependencies within the composite graph, we extend this framework to FKAGraphSAGE. We achieve this by replacing the traditional linear projection with our proposed FKANLayer (ϕ). This modification allows the model to process the combined neighborhood and central node features within a highly expressive spectral space. At the l-th layer, the feature update process is defined as follows:

$$h_{\mathcal{N}(v)}^{(l)} = \text{MEAN}\left(\{h_u^{(l-1)} \mid u \in \mathcal{N}(v)\}\right), \tag{21}$$

$$h_v^{(l)} = \phi\left(h_v^{(l-1)} \parallel h_{\mathcal{N}(v)}^{(l)}\right), \tag{22}$$

where MEAN$(\cdot)$ represents the mean pooling aggregator over the neighbor set $\mathcal{N}(v)$, and $[\ \|\ ]$ denotes the vector concatenation operation. The aggregated neighborhood context $h_{\mathcal{N}(v)}^{(l)}$ is concatenated with the feature of the central node $h_v^{(l-1)}$ and subsequently fed into the FKANLayer. By upgrading the feature transformation process with Fourier basis functions, FKAGraphSAGE efficiently learns deep, non-linear interactions between a node and its local context.

B.2 DATA RETRIEVAL AND PROCESSING DETAILS

To ensure the reproducibility of the experiments, we provide the detailed configurations for data retrieval and structural processing as follows:

AlphaFold Configuration. The 3D structures of target proteins were primarily retrieved from the AlphaFold Protein Structure Database (v6). For a small subset of sequences unavailable in this database, AlphaFold v2.3.2 was deployed locally to predict their structures, selecting the highest-ranked conformation (rank_0). To prevent highly disordered and unstructured regions from compromising the contact graphs, a confidence filtering mechanism was implemented. Specifically, residues with a predicted Local Distance Difference Test (pLDDT) score of less than 50 were excluded during the protein graph construction phase.

UniProt Retrieval Details. To systematically acquire experimentally validated binding sites (S_{uni}) necessary for the complementary binding site acquisition strategy, UniProtKB was accessed programmatically via its REST API. Structural and functional feature annotations were parsed using specific keys, retaining only those residues annotated with type="Binding site", type="Active site", or type="Site". Combining these experimentally grounded annotations with P2Rank predictions determines the scope of the binding pocket. This strategy reduces the effective protein sequence length, ensuring that the structural priors injected into the composite graph focus on interaction-critical intervals rather than noisy or redundant regions.

B.3 BASELINES

In this section, we detail the 18 state-of-the-art baselines used to evaluate the performance of our proposed SGFF-DTA on the Davis and KIBA datasets. These carefully selected methods cover a broad spectrum of computational strategies, including traditional machine learning, sequence-based deep learning, graph-based or multi-modal architectures, and recent cutting-edge models. A brief description of each baseline is provided below:

- **KronRLS** [Pahikkala et al., 2015]: A similarity-based machine learning method that employs the Kronecker Regularized Least Squares algorithm. It formulates the prediction of drug-target binding affinities by integrating 2D compound structural similarities and target protein sequence similarities into a unified mathematical framework.

- **SimBoost** [He et al., 2017]: A non-linear machine learning approach that utilizes gradient boosting regression trees. It constructs extensive network-based and similarity-based features from drugs, targets, and drug-target pairs to capture the entire interaction spectrum and predict continuous binding affinity values.
- **DeepCDA** [Abbasi et al., 2020]: A deep learning framework combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) layers to extract local substructure occurrence patterns. It specifically incorporates a two-sided attention mechanism and adversarial domain adaptation to effectively predict compound-protein affinities across different data domains.
- **DeepDTA** [Öztürk et al., 2018]: One of the pioneering deep learning models in this field, utilizing 1D Convolutional Neural Networks (CNNs) to extract local representation features directly from drug SMILES strings and protein sequences.
- **ML-DTI** [Yang et al., 2021]: An interpretable deep learning model that introduces a mutual learning mechanism using multi-head and position-aware attention. This global perspective approach dynamically bridges the gap between otherwise independent drug and target CNN encoders to enhance both prediction performance and interpretability.
- **DeepGLSTM** [Mukherjee et al., 2022]: A hybrid deep learning architecture that integrates Graph Convolutional Networks (GCNs) with LSTMs. It effectively captures the high-dimensional topological structures of drug molecular graphs alongside sequential target features to generate robust representations for affinity prediction.
- **DeepGS** [Lin et al., 2020]: A deep representation learning framework that extracts local chemical context and topological structure by encoding amino acids and SMILES sequences with advanced embedding techniques (Prot2Vec and Smi2Vec) and processing them through a combination of CNNs, GATs, and BiGRUs.
- **GraphDTA** [Nguyen et al., 2021]: A prominent deep learning architecture that represents drugs as molecular graphs rather than 1D strings, utilizing Graph Neural Networks (GNNs) to directly capture the chemical bonds among atoms alongside CNN-encoded protein sequences.
- **MGraphDTA** [Yang et al., 2022]: A deep multiscale graph neural network integrating a super-deep 27-layer GNN to capture both local and global compound structures simultaneously, featuring a visual explanation method (Grad-AAM) for chemical interpretability.
- **AttentionMGT-DTA** [Wu et al., 2024a]: A multi-modal prediction model that represents drugs as molecular graphs and targets as binding pocket graphs, employing graph transformers and dual attention mechanisms to seamlessly integrate and interact structural information.
- **MultiDTA** [Deng et al., 2024]: An end-to-end representation learning framework that utilizes four distinct channels—combining CNNs, LSTMs, and GCNs—to comprehensively extract sequential and spatial structural features, which are dynamically fused via an attention mechanism.
- **GLCN-DTA** [Qi et al., 2024]: An extended graph learning-convolutional network that innovatively learns a soft adjacency matrix to refine the contextual structures of drug and protein graphs, enabling the extraction of richer topological representations for DTA prediction.
- **LLMDTA** [Tang et al., 2025]: A novel cold-start prediction framework that leverages biological large language models (Mol2Vec for drugs and ESM2 for proteins) combined with a bilinear attention module to capture interactive molecular features and improve generalization on unseen data.
- **MambaTransDTA** [Lou et al., 2025]: An innovative hybrid sequence modeling architecture that integrates the linear-time, long-range dependency extraction capability of Mamba with the local interaction modeling of Transformers to comprehensively estimate binding affinity.
- **SMFF-DTA** [Wang et al., 2025]: A sequential multi-feature fusion method that utilizes multiple attention blocks to efficiently encode both the structural information and physicochemical properties of drugs and targets, avoiding the high computational overhead of complex molecular graphs.
- **MF-DTA** [Kang et al., 2025]: A multi-modal feature fusion model that innovatively introduces BRICS-based molecular fragment graphs to explicitly capture the structural characteristics and pharmacophore-related features of drug molecules for robust affinity prediction.
- **DMFF-DTA** [He et al., 2025]: A dual-modality neural network that seamlessly fuses sequence and graph information while addressing the scale disparity between drugs and proteins by utilizing AlphaFold2 to construct binding-site-focused contact maps.

- **MutualDTA** [Yuan et al., 2025]: An interpretable deep learning model that utilizes large-scale pretrained models and introduces a Mutual-Attention module to accurately capture the intermolecular interactions and binding sites between drug atoms and protein amino acid residues.