MultiGeo: Predicting Drug-Target Affinity via Adaptive Multi-Conformation Ensemble Learning

Ruida Zeng[1], **Cheng Guo**[1], **Yajie Meng**[2], **Xunkun Cheng**[1], **Zhiwei Xu**[1], **Xiangzheng Fu**[3], **Pan Zeng**[4], **Shuting Jin**[1*], **Junlin Xu**[1*]

[1]College of Computer Science and Technology, Wuhan University of Science and Technology
[2]School of Computer Science and Artificial Intelligence, Wuhan Textile University
[3]Computer Science and Artificial Intelligence, Shenzhen University of Advanced Technology
[4]College of Computer and Information Science, Chongqing Normal University
{202305403095, chengguo, xunkuncheng, xuzhiwei, shutingjin}@wust.edu.cn, {myj, xjl}@hnu.edu.cn, fuxiangzheng@suat-sz.edu.cn, zengpan@cqnu.edu.cn

Abstract

Predicting drug–target affinity (DTA) is central to drug discovery, yet most deep learning models rely on a single static protein structure, neglecting the conformational heterogeneity that underlies many binding mechanisms. We propose MultiGeo, a DTA prediction framework that explicitly leverages multiple protein conformations rather than a single snapshot. For each target, MultiGeo starts from an ensemble of structure predictions and uses confidence scores to adaptively select a small set of high-quality, diverse conformers. A hierarchical structural encoder then extracts multi-scale geometric features from these conformers, which are aggregated by a GRU to obtain an ensemble-level representation. To avoid indiscriminately mixing noisy or redundant views, we introduce a disagreement-aware gating mechanism that adaptively fuses this ensemble representation with the dominant structure only when the additional conformers provide complementary information. Finally, ligand–target interactions are modeled via a block-wise cross-attention module that captures multi-perspective dependencies between ligand and protein features. Extensive experiments on multiple DTA benchmarks demonstrate that MultiGeo consistently outperforms state-of-the-art baselines, showing that explicitly encoding conformational diversity yields more accurate and robust affinity prediction.

1 Introduction

Drug-Target Affinity (DTA) prediction, quantifying the binding strength between a ligand and a protein, represents a central and persistent challenge in early-stage drug discovery [Vefghi *et al.*, 2025]. While modern high-throughput screening (HTS) technologies [Shaik and Balya, 2025] offer high fidelity, their immense financial and temporal burdens render the exhaustive exploration of vast chemical libraries impractical. To mitigate these experimental costs, accurate computational prediction has become indispensable for accelerating virtual screening [Ashburn and Thor, 2004]. Consequently, data-driven deep learning models have emerged as a powerful alternative.

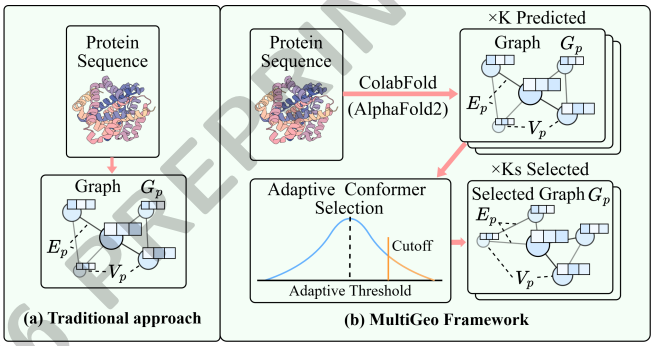

Figure 1: Comparison of protein graph representation frameworks. (a) The traditional approach utilizing a single static input. (b) Our MultiGeo framework incorporating an **Adaptive Conformer Selection** strategy for ensemble representation.

Early sequence-based models, such as DeepDTA [Öztürk *et al.*, 2018] and WideDTA [Öztürk *et al.*, 2019], utilized CNNs to process SMILES and amino acid sequences. To better capture long-range dependencies within these sequences, subsequent models integrated RNNs [Karimi *et al.*, 2019] and LSTM [Zheng *et al.*, 2020] networks. However, these sequence-centric models were inherently limited in their capacity to model the true structural environment and spatial organization that govern molecular recognition and binding. This limitation prompted a shift toward graph-based representations, which excel at modeling topological structure. Pioneering works like GraphDTA [Nguyen *et al.*, 2021] introduced graph neural networks (GNNs) to encode the atomic graph of drug molecules. This paradigm was later extended to proteins, with models such as DGraphDTA [Jiang *et al.*, 2020] and WGNN-DTA [Jiang *et al.*, 2022] constructing residue-level interaction graphs from predicted contact maps,

*Corresponding authors.

thereby incorporating a coarse form of structural information. Furthermore, the integration of Graph Transformers, as seen in AttentionMGT-DTA [Wu *et al.*, 2024], has refined this approach by capturing intricate non-local dependencies, particularly within key regions like binding pockets. More recently, advances in structural biology, especially AlphaFold2 [Jumper *et al.*, 2021], have accelerated structure-aware DTA modeling. Some approaches like CASTER-DTA [Kumar *et al.*, 2025] directly leverage equivariant graph neural networks to capture 3D spatial arrangements.

Despite these significant advancements, a critical limitation remains: most models rely on static protein structures, ignoring the conformational flexibility required for binding [Vefghi *et al.*, 2025]. As illustrated in Figure 1(a), whether relying on 1D sequences, advanced Graph Transformers, or 3D complexes, current models almost universally treat proteins as rigid, static entities, dependent on a single crystal structure or a single computational prediction. This simplification starkly contrasts with biological reality, where proteins exist as dynamic conformational ensembles [Henzler-Wildman and Kern, 2007]. By neglecting the intrinsic structural plasticity and conformational heterogeneity [Kannan and Naganathan, 2025], these single-snapshot models often fail to capture the full spectrum of interaction possibilities, limiting their generalization to novel targets.

To bridge this gap, we introduce MultiGeo, a novel framework designed to shift the paradigm from single-snapshot modeling to multi-geometry aggregation (as shown in Figure 1b). Instead of relying on a single structure, our model leverages an energy-ranked ensemble of conformers predicted by AlphaFold2 to explicitly encode structural plasticity [Kalakoti and Wallner, 2025]. By strictly distinguishing between informative structural variance and random prediction noise, MultiGeo transforms raw conformational diversity into robust predictive signals.

Specific contributions of this work are as follows:

- We propose MultiGeo, an end-to-end framework that integrates an adaptive conformer selection strategy. This approach utilizes confidence scores to filter the AlphaFold2 ensemble.
- Our framework uniquely integrates a Hierarchical Structural Encoder with a disagreement-aware gating mechanism, enabling the adaptive fusion of ensemble plasticity with the dominant structural representation.
- Extensive experiments on two DTA datasets demonstrate the viability, effectiveness, and superiority of our MultiGeo.

2 Materials and methods

2.1 Geometric Graph Representation

Drug Graph Representation

As shown in Figure 2(a), to effectively capture the chemical topology and physicochemical properties of ligands, each drug molecule is represented as a molecular graph $G_d = (V_d, E_d)$. Given a drug SMILES string, we utilize the built-in algorithm of RDKit to generate the conformer of the molecule. The set of nodes V_d corresponds to the heavy atoms, while the set of edges E_d comprises covalent chemical bonds. Each node $v \in V_d$ is associated with a compact feature vector $\mathbf{x}_v \in R^{78}$. The node features encompass a selected set of atomic attributes, including atomic symbol, degree, total number of hydrogens, implicit valence, and aromaticity. Similarly, edges are initialized with feature vectors $\mathbf{x}_e \in R^{10}$ encoding bond type, conjugation status, ring membership, and stereochemistry. This topological graph structure allows the model to learn from the chemical connectivity of the ligand.

Protein Graph Representation

Regarding the target protein, Figure 2(a) depicts the generation of a raw ensemble consisting of K_{total} predicted structures via ColabFold (AlphaFold2). Unlike previous methods that construct graphs solely based on Euclidean distances from a single snapshot, we propose a confidence-aware graph construction strategy utilizing the outputs from the protein language model. Each protein conformer is modeled as a geometric graph $G_p = (V_p, E_p)$, where nodes represent amino acid residues centered at C_α coordinates. The edges are constructed based on the Predicted Aligned Error (PAE) matrix. An edge is established between residue i and j if the expected error in their relative position is below a strict threshold (e.g., 8.0Å). This topological structure filters out unreliable contacts, ensuring that the graph edges represent high-confidence physical interactions.

2.2 MultiGeo Model Architecture

Adaptive Conformer Selection Strategy

To filter prediction noise, we apply an Adaptive Conformer Selection step shown in Figure 2(a), which utilizes the pLDDT (predicted Local Distance Difference Test) score as the metric for structural reliability. Instead of a fixed hard threshold, we employ an adaptive filtering strategy based on the statistical distribution of the entire dataset. First, we compute the global standard deviation (σ_{global}) of pLDDT scores across all conformers in the training set. For a specific target protein t with a maximum confidence score $S_{max}^{(t)}$ (typically from the Rank 1 structure), we define a dynamic acceptance threshold $\tau^{(t)}$:

$$\tau^{(t)} = S_{max}^{(t)} - \lambda \cdot \sigma_{global} \tag{1}$$

where λ is a hyperparameter controlling the tolerance (e.g., $\lambda = 1.0$). We construct the final ensemble $\mathcal{G}_p$ by retaining only those conformers whose mean pLDDT scores exceed $\tau^{(t)}$, resulting in a refined set of K_s selected conformers (where $1 \leq K_s \leq K_{total}$). This strategy ensures that the model includes alternative conformers only when their structural plausibility is statistically comparable to the dominant state, thereby effectively filtering out low-quality predictions while preserving valid conformational diversity.

Ensemble Landscape Feature Extraction

Upon determining the set of K_s valid conformers via the adaptive selection strategy, we construct an ordered feature sequence by extracting high-level representations from the shared Hierarchical Structural Encoder. Specifically, let $\mathcal{S}_{prot}$ denote the sequence of graph-level representations extracted

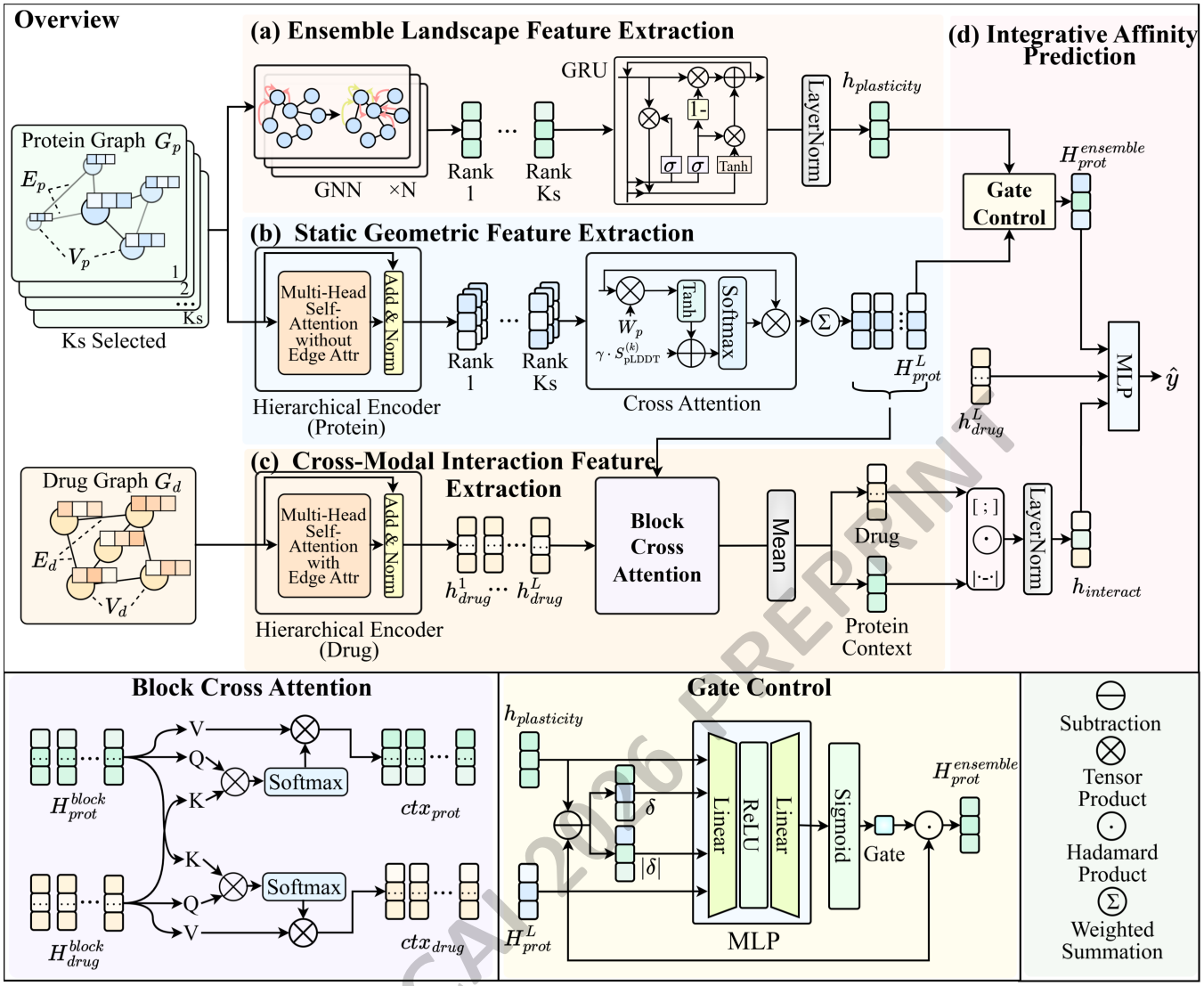

Figure 2: Overview of the MultiGeo framework. The architecture consists of four key modules: (a) **Ensemble Landscape Feature Extraction** uses a GRU to capture structural plasticity. (b) **Static Geometric Feature Extraction** derives a robust static reference via cross-attention pooling. (c) **Cross-Modal Interaction Feature Extraction** captures ligand-target dependencies through block-wise cross-attention. (d) **Integrative Affinity Prediction** adaptively fuses these features using a **Disagreement-Aware Gate Control** mechanism to predict the final binding affinity.

from the final block L for each selected conformer (sorted by confidence):

$$\mathcal{S}_{prot} = \left[\mathbf{H}^{(1)}_{prot}, \mathbf{H}^{(2)}_{prot}, \ldots, \mathbf{H}^{(K_s)}_{prot}\right] \tag{2}$$

where $\mathbf{H}^{(k)}_{prot} \in R^d$ represents the geometric embedding of the k-th conformer (for $k = 1, \ldots, K_s$). This sequence captures the variation of structural features from the most probable state (Rank 1) to alternative meta-stable states.

To synthesize these distinct snapshots into a holistic representation, we employ a Unidirectional Gated Recurrent Unit (GRU) to sequentially process the conformer features and learn the latent dependencies between different energy states:

$$\mathbf{h}_k = \text{GRU}\left(\mathbf{H}^{(k)}_{prot}, \mathbf{h}_{k-1}\right) \tag{3}$$

The final hidden state $\mathbf{h}_{K_s}$, obtained after processing the entire sequence, serves as the compact representation of the ensemble.

We define the raw plasticity context, denoted as $\mathbf{h}_{plasticity}$, by projecting this final accumulated state:

$$\mathbf{h}_{plasticity} = \text{LayerNorm}\left(\mathbf{W}_{ens}\mathbf{h}_{K_s} + \mathbf{b}_{ens}\right) \tag{4}$$

This vector $\mathbf{h}_{plasticity}$ encapsulates the inherent latent potential for deformation of the target protein, providing the necessary input for the subsequent gating mechanism.

Static Geometric Feature Extraction

To extract robust static representations for both the ligand and the target, we design a unified Hierarchical Structural Encoder composed of L stacked blocks. This encoder is

shared across the drug graph G_d and each of the selected protein conformer graphs $\{G_p^{(1)}, \ldots, G_p^{(K_s)}\}$. Within each structural block, feature extraction commences with a Transformer Convolution layer to dynamically aggregate neighbor information. For a node i at layer l, the updated feature vector $\mathbf{h}_i'$ is computed as:

$$\mathbf{h}_i' = \mathbf{W}_O\Big(\mathbf{h}_i^{(l-1)} + \sum_{j\in\mathcal{N}(i)} \alpha_{ij}\big(\mathbf{W}_V\mathbf{h}_j^{(l-1)} + \mathbf{W}_E\mathbf{e}_{ij}\big)\Big) \quad (5)$$

where $\mathbf{e}_{ij}$ denotes edge features (specifically integrated for the ligand to capture chemical bond attributes, while set to zero for the protein), and α_{ij} is the attention coefficient. This is followed by an Intra-Graph Attention layer to capture long-range dependencies. Crucially, each block concludes with a Self-Attention Graph Pooling (SAGPooling) operation to achieve hierarchical abstraction. A learnable importance score s_i is computed to adaptively weight nodes based on their structural significance:

$$s_i = \sigma(\mathbf{w}_{att}^T \cdot \text{GCN}(\mathbf{h}_i')), \quad \mathbf{H}^{(l)} = \text{Readout}(\mathbf{h}' \odot s) \quad (6)$$

While this process directly yields the hierarchical block features $\mathbf{H}_{drug}^{block}$ for the ligand, the protein requires synthesizing K_s conformers into a single stable reference. Therefore, we aggregate the protein outputs at each block level using a Cross-Attention Pooling mechanism. To prioritize high-quality structures, we incorporate the pLDDT score as an explicit bias. The attention score $s_k^{(l)}$ for conformer k at block l is computed as:

$$s_k^{(l)} = \mathbf{w}_{pool}^T \tanh(\mathbf{W}_p\mathbf{H}_{prot,k}^{(l)}) + \gamma \cdot \tilde{S}_{pLDDT}^{(k)} \quad (7)$$

where $\tilde{S}_{pLDDT}^{(k)}$ represents the normalized and centered confidence score, serving as a quality bias, and γ is a pre-defined weighting factor controlling its influence. Based on the normalized weights $\alpha_k^{(l)}$ from a softmax function, we compute the weighted sum:

$$\mathbf{H}_{prot}^{fused,(l)} = \sum_{k=1}^{K_s} \alpha_k^{(l)}\mathbf{H}_{prot,k}^{(l)} \quad (8)$$

The output from the final block L, denoted as $\mathbf{H}_{prot}^{L}$, serves as the static dominant representation, acting as the consensus baseline for prediction.

Cross-Modal Interaction Feature Extraction

To rigorously evaluate the compatibility between the ligand and the target, we implement a deep interaction mechanism based on Block-Wise Cross-Attention. Unlike conventional methods that rely on the late fusion of global embeddings, this approach enables the ligand and target to interact at multiple structural resolutions, ranging from local atomic environments to global domain-level motifs.

Let $\mathbf{H}_{drug}^{block}$ denote the output of the Hierarchical Structural Encoder for the drug, and let $\mathbf{H}_{prot}^{fused}$ denote the Fused Hierarchical Block Features for the protein derived in Section 3.2.2:

$$\mathbf{H}_{drug}^{block} \in R^{L\times d}, \quad \mathbf{H}_{prot}^{fused} \in R^{L\times d} \quad (9)$$

where L is the number of structural blocks and d is the hidden dimension.

To model the mutual influence between the molecules, we first compute Bidirectional Block Attention. The drug features query the protein features to form a drug-centric context, while the protein features symmetrically query the drug features. We project $\mathbf{H}_{drug}^{block}$ and $\mathbf{H}_{prot}^{fused}$ into queries ($\mathbf{Q}$), keys ($\mathbf{K}$), and values ($\mathbf{V}$). The drug-to-protein context $\mathbf{C}_{d\to p}$ and protein-to-drug context $\mathbf{C}_{p\to d}$ are computed as:

$$\begin{aligned}\mathbf{C}_{d\to p} &= \text{softmax}\left(\frac{\mathbf{Q}_d\mathbf{K}_p^T}{\sqrt{d}}\right)\mathbf{V}_p \\ \mathbf{C}_{p\to d} &= \text{softmax}\left(\frac{\mathbf{Q}_p\mathbf{K}_d^T}{\sqrt{d}}\right)\mathbf{V}_d\end{aligned} \quad (10)$$

To aggregate these multi-scale interactions into unified vector representations, we apply mean pooling across the block dimension L, augmented with residual connections to the original features:

$$\begin{aligned}\bar{\mathbf{c}}_d &= \text{Mean}(\mathbf{C}_{d\to p} + \mathbf{H}_{drug}^{block}) \\ \bar{\mathbf{c}}_p &= \text{Mean}(\mathbf{C}_{p\to d} + \mathbf{H}_{prot}^{fused})\end{aligned} \quad (11)$$

Integrative Affinity Prediction

This final stage integrates the extracted feature streams to predict the binding affinity. First, building upon the global drug and protein contexts ($\bar{\mathbf{c}}_d$ and $\bar{\mathbf{c}}_p$) derived in the previous section, we move beyond simple feature concatenation to rigorously evaluate compatibility. We employ a multi-perspective interaction fusion strategy. Inspired by heuristic matching methods in natural language inference, this operation explicitly decomposes the interaction into distinct views, utilizing the element-wise product to capture similarity and the absolute difference to measure discrepancy:

$$\mathbf{h}_{interact} = \mathbf{W}_{fuse}[\bar{\mathbf{c}}_d; \bar{\mathbf{c}}_p; \bar{\mathbf{c}}_d \odot \bar{\mathbf{c}}_p; |\bar{\mathbf{c}}_d - \bar{\mathbf{c}}_p|] + \mathbf{b}_{fuse} \quad (12)$$

where $\odot$ represents the element-wise Hadamard product, $|\cdot|$ denotes absolute difference, and $[\cdot;\cdot]$ denotes concatenation. $\mathbf{h}_{interact}$ serves as the robust Cross-Modal Interaction Representation.

Simultaneously, to adaptively modulate the contribution of structural plasticity, we process the raw plasticity context $\mathbf{h}_{plasticity}$ (from Module a) using a disagreement-aware gating mechanism. Let $\mathbf{H}_{prot}^{L}$ (from Module b) be the static reference of the dominant state. The gating signal g is computed based on the heuristic comparison between the dynamic context and the static reference:

$$\boldsymbol{\delta} = \mathbf{h}_{plasticity} - \mathbf{H}_{prot}^{L} \quad (13)$$

$$g = \sigma\left(\mathbf{W}_{gate}[\mathbf{h}_{plasticity}\,;\,\mathbf{H}_{prot}^{L}\,;\,\boldsymbol{\delta}\,;\,|\boldsymbol{\delta}|]\right) \quad (14)$$

This mechanism allows the model to emphasize the ensemble landscape features only when the conformational landscape provides significant complementary information to the static structure. The gated ensemble feature is formally defined as:

$$\mathbf{H}_{prot}^{ensemble} = \beta \cdot g \odot \mathbf{h}_{plasticity} \quad (15)$$

where β is a learnable scaling factor initialized to 0.

Dataset	Metric	Graph(GAT)	Graph(GIN)	WGNN(GAT)	WGNN(GCN)	G-K Bert	DeepDTA	DeepDTAGen	GTOmega	MultiGeo
DAVIS	MSE (↓)	0.591	0.614	0.639	0.751	0.547	0.507	0.662	<u>0.479</u>	**0.454**
	RMSE (↓)	0.769	0.783	0.799	0.867	0.740	0.712	0.814	<u>0.692</u>	**0.674**
	PCC(↑)	0.314	0.220	0.220	0.285	0.332	0.337	0.194	<u>0.424</u>	**0.449**
	r_m^2 (↑)	0.094	0.040	0.042	0.066	0.098	0.103	0.032	<u>0.154</u>	**0.171**
KIBA	MSE (↓)	0.522	0.418	0.460	0.498	0.405	0.494	0.431	<u>0.384</u>	**0.365**
	RMSE (↓)	0.722	0.647	0.678	0.705	0.636	0.703	0.656	<u>0.619</u>	**0.604**
	PCC (↑)	0.522	0.666	0.611	0.565	0.666	0.577	0.646	<u>0.686</u>	**0.690**
	r_m^2 (↑)	0.268	0.405	0.339	0.296	0.408	0.294	0.370	<u>0.418</u>	**0.460**

The best score is in **bold**, the second best is <u>underlined</u>.

Table 1: Performance comparison on Davis and KIBA datasets (Cold-Drug setting).

Dataset	Metric	Graph(GAT)	Graph(GIN)	WGNN(GAT)	WGNN(GCN)	G-K Bert	DeepDTA	DeepDTAGen	GTOmega	MultiGeo
DAVIS	MSE (↓)	0.568	0.557	0.572	0.601	<u>0.554</u>	0.571	0.662	0.556	**0.455**
	RMSE (↓)	0.753	0.746	0.756	0.775	<u>0.744</u>	0.756	0.814	0.746	**0.674**
	PCC (↑)	0.314	0.451	0.404	0.377	<u>0.457</u>	0.083	0.194	0.454	**0.593**
	r_m^2 (↑)	0.094	0.186	0.162	0.136	<u>0.191</u>	0.006	0.032	0.189	**0.311**
KIBA	MSE (↓)	0.574	0.507	0.446	0.448	0.488	0.444	0.500	<u>0.417</u>	**0.389**
	RMSE (↓)	0.757	0.712	0.667	0.669	0.698	0.667	0.707	<u>0.645</u>	**0.624**
	PCC (↑)	0.432	0.526	0.611	0.607	0.551	0.628	0.534	<u>0.648</u>	**0.677**
	r_m^2 (↑)	0.180	0.176	0.343	0.339	0.397	0.387	0.284	<u>0.397</u>	**0.404**

The best score is in **bold**, the second best is <u>underlined</u>.

Table 2: Performance comparison on Davis and KIBA datasets (Cold-Target setting).

Finally, we construct the comprehensive feature vector by concatenating the static drug representation, the gated ensemble feature, and the interaction feature:

$$\mathbf{h}_{final} = \left[\mathbf{h}_{drug}^{L}\,;\,\mathbf{H}_{prot}^{ensemble}\,;\,\mathbf{h}_{interact}\right] \tag{16}$$

This vector is then passed through an MLP regressor to yield the predicted binding affinity $\hat{y}$:

$$\hat{y} = \text{MLP}_{pred}(\mathbf{h}_{final}) \tag{17}$$

To train the MultiGeo framework end-to-end, we utilize the Mean Squared Error (MSE) as the loss function:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{18}$$

This strategy ensures that the final prediction is informed by local geometry, deep cross-modal dependencies, and the global conformational plasticity of the target.

3 Experiments and Results

3.1 Experimental Setup

Dataset. We evaluated our MultiGeo model on two benchmark datasets: Davis [Davis *et al.*, 2011] and KIBA [Tang *et al.*, 2014]. Binding affinity values are typically expressed by dissociation constants (K_d), inhibition constants (K_i), or half maximal inhibitory concentrations (IC_{50}). For the Davis dataset, affinities are K_d values, converted to the logarithmic scale as $pK_d = -\log_{10}(K_d/10^9)$, ranging from 5.0 to 10.8. The KIBA dataset uses a unified scoring system based on K_d, K_i, and IC_{50}, with values mainly between 10.0 and 17.2. Duplicate samples were removed to ensure data quality.

To rigorously assess generalizability, we employed three challenging splitting settings, consistent with the Therapeutics Data Commons benchmarks:

- Drug Cold-start: Test set drugs are unseen during training, while targets are shared. This evaluates the model's ability to extrapolate to novel compounds.
- Target Cold-start: Test set targets are new, but drugs are seen during training. This assesses generalization to unknown proteins.
- Drug-Target Cold-start: Both drugs and targets in the test set are entirely novel. This represents the most challenging and realistic scenario for DTA prediction.

To ensure comprehensive comparability with existing literature, we additionally conducted standard 5-fold cross-validation experiments. Detailed results and implementation codes are available at https://github.com/refasdas434-tech/MultiGeo.

Evaluation Metrics. To quantitatively evaluate the predictive performance, we employed four standard metrics: Mean Squared Error (MSE), Root Mean Square Error (RMSE), Pearson Correlation Coefficient (PCC), and the squared correlation coefficient with zero intercept (r_m^2). For MSE and RMSE, lower values indicate superior performance, whereas for PCC and r_m^2, higher values imply better predictive accuracy.

Dataset	Metric	Graph(GAT)	Graph(GIN)	WGNN(GAT)	WGNN(GCN)	G-K Bert	DeepDTA	DeepDTAGen	GTOmega	MultiGeo
DAVIS	MSE (↓)	0.825	0.531	0.874	0.663	0.548	0.553	0.632	0.522	**0.494**
	RMSE (↓)	0.908	0.729	0.935	0.815	0.740	0.743	0.795	0.722	**0.703**
	PCC (↑)	0.076	0.030	0.026	0.070	0.028	0.104	0.075	0.160	**0.266**
	r_m^2 (↑)	0.004	0.001	0.003	0.004	0.001	0.010	0.006	0.025	**0.069**
KIBA	MSE (↓)	0.724	0.753	0.662	0.701	0.678	0.657	0.653	0.601	**0.596**
	RMSE (↓)	0.851	0.868	0.813	0.837	0.824	0.811	0.808	0.775	**0.772**
	PCC (↑)	0.379	0.342	0.383	0.343	0.338	0.410	0.372	0.460	**0.472**
	r_m^2 (↑)	0.129	0.109	0.136	0.108	0.112	0.151	0.136	0.199	**0.208**

The best score is in **bold**, the second best is underlined.

Table 3: Performance comparison on Davis and KIBA datasets (Cold-Drug-Target setting).

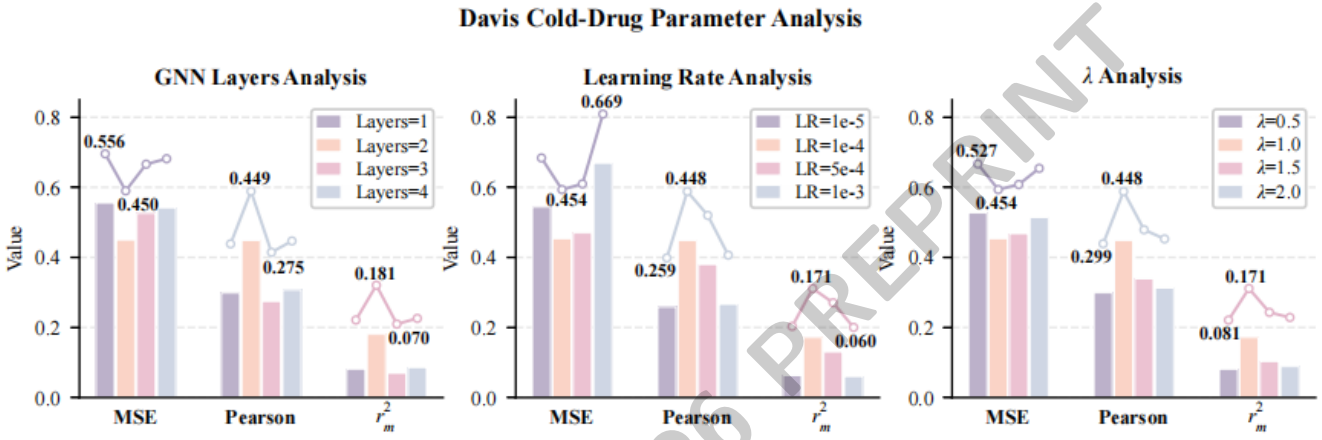

Figure 3: Parameter sensitivity analysis on the Davis dataset under the Cold-Drug setting, including the effects of GNN layers, learning rate, and the standard deviation factor (λ).

Baselines. To validate the efficacy of our **MultiGeo**, we evaluate it against three state-of-the-art sequence-based techniques: **DeepDTA** [Öztürk *et al.*, 2018], **DeepDTAGen** [Shah *et al.*, 2025], and **G-K BertDTA** [Qiu *et al.*, 2024]. Additionally, we perform experiments with three structure-based techniques: **GraphDTA** [Nguyen *et al.*, 2021] (incorporating GAT_GCN and GIN variants), **WGNN-DTA** [Jiang *et al.*, 2022] (incorporating GAT and GCN variants), and **DTA-GTOmega** [Quan *et al.*, 2025].

Implementation Details. The proposed MultiGeo is implemented utilizing PyTorch [Paszke *et al.*, 2019] and the Deep Graph Library (DGL) [Wang, 2019]. All experiments were executed on a workstation equipped with an NVIDIA GeForce RTX 4090 GPU to accelerate the training process.

3.2 Comparison with Baselines

Tables 1-3 report the performances of the proposed method and recent state-of-the-art approaches on the DAVIS and KIBA datasets under three challenging scenarios. Experimental results demonstrate that the proposed MultiGeo comprehensively outperforms other models on the DAVIS dataset when faced with unseen targets or drugs, which verifies the effectiveness and generalization of our framework. In three challenging scenarios, our method achieves PCC gains of 5.9% (Cold-Drug), 30.6% (Cold-Target), and a remarkable 66.2% (Cold-Drug-Target) over the best baseline models. This signifies that our MultiGeo is capable of capturing subtle yet critical patterns in drug-target interactions, specifically the structural plasticity that other static models may overlook. Besides, it can be observed that our MultiGeo also achieves promising performance on the KIBA dataset. In comparison with the runner-up GTOmega, MultiGeo consistently maintains the lowest MSE across all settings. These observations further verify the robustness of our model across different data distributions. Among all baselines, the performance of structure-based methods (e.g., GTOmega, WGNN-DTA) is generally better than that of sequence-based methods (e.g., DeepDTA, DeepDTAGen). This indicates that simply extracting representations from target sequences is not sufficient to capture the inherent properties, and the spatial arrangement and three-dimensional conformation of proteins play a crucial role in their affinity to drugs. In comparison, the proposed method not only leverages high-fidelity geometric graphs but also exploits the conformational landscape in a hierarchical manner and adaptively fuses them via a disagreement-aware gating mechanism for end-to-end learning. This approach allows for a more comprehensive understanding of the drug-target interaction landscape, leading to improved predictive

accuracy.

3.3 Ablation study

Effectiveness of designed components. We conduct an ablation analysis by evaluating the following MultiGeo variants:

- **w/o AttnFusion:** Replaces attention pooling with the single Rank-1 structure, discarding ensemble diversity.
- **w/o DrugRepr:** Removes the static drug representation ($\mathbf{h}_{drug}^{L}$) from the final concatenation.
- **w/o GRU:** Removes the GRU module, thereby neglecting the structural plasticity context.
- **w/o Gating:** Removes the gating mechanism, directly concatenating unfiltered ensemble features with other representations.
- **w/o Interaction:** Excludes the explicit interaction features ($\mathbf{h}_{interact}$) from the final prediction.

Methods	Davis			KIBA		
	MSE↓	PCC↑	r_m^2↑	MSE↓	PCC↑	r_m^2↑
MultiGeo	**0.454**	**0.449**	**0.171**	**0.365**	**0.690**	**0.460**
w/o AttnFusion	0.561	0.196	0.036	0.419	0.634	0.345
w/o DrugRepr	0.503	0.356	0.112	0.417	0.669	0.376
w/o GRU	0.552	0.178	0.031	0.406	0.673	0.403
w/o Gating	0.718	0.250	0.043	0.418	0.661	0.381
w/o Interaction	0.549	0.216	0.046	0.501	0.550	0.298

Table 4: Ablation study on Davis and KIBA dataset under the Cold-Drug setting

Table 4 reports the performance of MultiGeo and its five variants on the Davis and KIBA datasets. It can be observed that all variants produce decreased performances compared to the full model, verifying that each designed component contributes positively to the final prediction. Specifically: (1) The variant w/o Gating exhibits the most severe performance degradation on Davis (MSE rises to 0.718). This demonstrates that indiscriminately concatenating ensemble features introduces substantial noise, thereby validating the necessity of our disagreement-aware gating mechanism to filter redundancy and retain only complementary signals. (2) The superiority of MultiGeo over w/o GRU and w/o AttnFusion confirms that hierarchically integrating the structural plasticity context with a robust static consensus provides a more precise description of the protein than single-snapshot approaches. (3) The notable decline in w/o Interaction highlights the effectiveness of the multi-perspective fusion strategy in capturing ligand-target compatibility. (4) The reduced performance of w/o DrugRepr indicates that preserving the fundamental static representation remains essential. Collectively, these results confirm that MultiGeo effectively creates a synergy between static geometry and conformational plasticity.

3.4 Hyper-parameter Sensitivity Analysis

Effect of GNN layers and learning rate. We first evaluate the impact of network depth and learning rate (LR). As shown in Figure 3 (Left), the model achieves optimal performance with 2 GNN layers. Increasing the depth beyond this leads to performance degradation, likely due to the over-smoothing phenomenon where node features become indistinguishable. Regarding the learning rate (Figure 3 Middle), the best result is observed at $1e$-4. A larger LR ($1e$-3) causes severe instability and poor convergence (MSE surges to 0.6691), while a smaller LR ($1e$-5) fails to escape local minima efficiently.

Effect of the standard deviation factor λ. Crucially, we investigate the coefficient λ in the adaptive conformer selection strategy to evaluate the trade-off between ensemble diversity and structural quality. As depicted in Figure 3 (Right), the model peaks when $\lambda = 1.0$. The results indicate that a higher λ (e.g., 2.0) introduces excessive noise from low-quality conformers, hindering effective learning. Conversely, a lower λ (e.g., 0.5) overly restricts the ensemble, filtering out valuable conformational variations and reducing the model to a near-static representation. Thus, $\lambda = 1.0$ strikes the optimal balance for capturing valid structural variations.

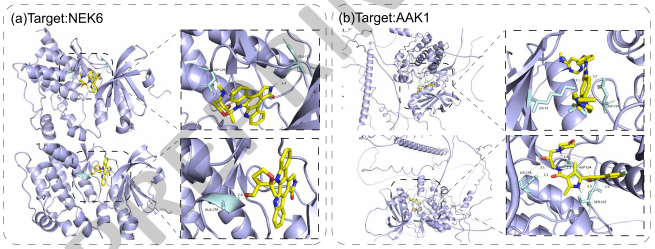

Figure 4: **Visualization of protein-ligand interactions.** (a) The binding poses of **NEK6** with Sunitinib (top) and Sorafenib (bottom). (b) The binding poses of **AAK1** with Tozasertib (top) and Danusertib (bottom). Key residues forming interactions are highlighted.

3.5 Case Study

To assess whether MultiGeo captures meaningful geometric interactions, we visually analyzed top-ranked unobserved drug–target pairs from the Davis dataset. As shown in Figure 4, the predicted complexes of NEK6 and AAK1 place ligands in plausible binding pockets and highlight interactions with key residues. These observations support the reliability of the predicted affinities.

4 Conclusion

In this work, we propose MultiGeo, a novel framework that shifts the paradigm of DTA prediction from single-snapshot modeling to conformational landscape aggregation. Addressing the limitations of rigid representations, MultiGeo explicitly encodes the intrinsic structural plasticity essential for complex binding mechanisms. By integrating the adaptive conformer selection strategy with the disagreement-aware gating mechanism, our framework rigorously filters prediction noise while intelligently fusing complementary plasticity features. Comprehensive experiments demonstrate that MultiGeo not only establishes a new state-of-the-art across challenging cold-start scenarios but also validates the critical importance of integrating multi-conformational ensembles to overcome the inherent rigidity prevalent in traditional approaches.

References

[Ashburn and Thor, 2004] Ted T Ashburn and Karl B Thor. Drug repositioning: identifying and developing new uses for existing drugs. *Nature reviews Drug discovery*, 3(8):673–683, 2004.

[Davis *et al.*, 2011] 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.

[Henzler-Wildman and Kern, 2007] Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, 2007.

[Jiang *et al.*, 2020] Mingjian Jiang, Zhen Li, Shugang Zhang, Shuang Wang, Xiaofeng Wang, Qing Yuan, and Zhiqiang Wei. Drug–target affinity prediction using graph neural network and contact maps. *RSC advances*, 10(35):20701–20712, 2020.

[Jiang *et al.*, 2022] Mingjian Jiang, Shuang Wang, Shugang Zhang, Wei Zhou, Yuanyuan Zhang, and Zhen Li. Sequence-based drug-target affinity prediction using weighted graph neural networks. *BMC genomics*, 23(1):449, 2022.

[Jumper *et al.*, 2021] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

[Kalakoti and Wallner, 2025] Yogesh Kalakoti and Björn Wallner. Afsample2 predicts multiple conformations and ensembles with alphafold2. *Communications Biology*, 8(1):373, 2025.

[Kannan and Naganathan, 2025] Adithi Kannan and Athi N Naganathan. Mapping the conformational heterogeneity intrinsic to the protein native ensemble. *Biochemistry*, 2025.

[Karimi *et al.*, 2019] Mostafa Karimi, Di Wu, Zhangyang Wang, and Yang Shen. Deepaffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics*, 35(18):3329–3338, 2019.

[Kumar *et al.*, 2025] Rachit Kumar, Joseph D Romano, and Marylyn D Ritchie. Caster-dta: equivariant graph neural networks for predicting drug–target affinity. *Briefings in Bioinformatics*, 26(5):bbaf554, 2025.

[Nguyen *et al.*, 2021] 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.

[Öztürk *et al.*, 2018] Hakime Öztürk, Arzucan Özgür, and Elif Ozkirimli. Deepdta: deep drug–target binding affinity prediction. *Bioinformatics*, 34(17):i821–i829, 2018.

[Öztürk *et al.*, 2019] Hakime Öztürk, Elif Ozkirimli, and Arzucan Özgür. Widedta: prediction of drug-target binding affinity. *arXiv preprint arXiv:1902.04166*, 2019.

[Paszke *et al.*, 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[Qiu *et al.*, 2024] Xihe Qiu, Haoyu Wang, Xiaoyu Tan, and Zhijun Fang. Gk bertdta: a graph representation learning and semantic embedding-based framework for drug-target affinity prediction. *Computers in Biology and Medicine*, 173:108376, 2024.

[Quan *et al.*, 2025] Lijun Quan, Jian Wu, Yelu Jiang, Deng Pan, and Lyu Qiang. Dta-gtomega: Enhancing drug-target binding affinity prediction with graph transformers using omegafold protein structures. *Journal of Molecular Biology*, 437(6):168843, 2025.

[Shah *et al.*, 2025] Pir Masoom Shah, Huimin Zhu, Zhangli Lu, Kaili Wang, Jing Tang, and Min Li. Deepdtagen: a multitask deep learning framework for drug-target affinity prediction and target-aware drugs generation. *Nature Communications*, 16(1):5021, 2025.

[Shaik and Balya, 2025] Neelufar Shama Shaik and Harika Balya. High-throughput computational screening for lead discovery and development. In *Advances in Pharmacology*, volume 103, pages 185–207. Elsevier, 2025.

[Tang *et al.*, 2014] 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.

[Vefghi *et al.*, 2025] Ali Vefghi, Zahed Rahmati, and Mohammad Akbari. Drug-target interaction/affinity prediction: Deep learning models and advances review. *Computers in Biology and Medicine*, 196:110438, 2025.

[Wang, 2019] Minjie Yu Wang. Deep graph library: Towards efficient and scalable deep learning on graphs. In *ICLR workshop on representation learning on graphs and manifolds*, 2019.

[Wu *et al.*, 2024] 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, 2024.

[Zheng *et al.*, 2020] Shuangjia Zheng, Yongjian Li, Sheng Chen, Jun Xu, and Yuedong Yang. Predicting drug–protein interaction using quasi-visual question answering system. *Nature Machine Intelligence*, 2(2):134–140, 2020.