# PAFDTA: PROGRESSIVE ADAPTIVE FUSION OF HETEROGENEOUS REPRESENTATIONS WITH CROSS-BRANCH CONTRASTIVE ALIGNMENT FOR DRUG-TARGET AFFINITY PREDICTION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Drug-target affinity (DTA) prediction depends on learning representations that preserve complementary information from molecular structure, protein sequence, pair-specific interactions, and relational context. We propose a heterogeneous representation-learning framework, namely PAFDTA, combining graph-based molecular features, pretrained protein representations, drug-target pair context, and topology-aware relational priors. Multi-scale Local-Global Aggregation captures complementary molecular patterns, while Cross-Branch Contrastive Alignment uses relational context to regularize the molecular latent representation before fusion. Progressive Adaptive Fusion then coordinates the molecular, protein, and relational representations through pair-conditioned weighting and relation-level interaction modeling, while the joint drug-target representation is retained as a direct pair-context pathway for affinity regression. We evaluate PAFDTA on Davis and KIBA using five-fold warm evaluation, component ablations, and cold-start settings with unseen drugs and/or targets. Additional analyses examine prediction consistency, uncertainty-aware candidate prioritization, structural interpretation of learned protein representations, and a BRAF-oriented repositioning case. The results support heterogeneous representation learning and adaptive representation integration as effective principles for DTA prediction.

## 1 INTRODUCTION

Drug discovery is often constrained by long development cycles, high development costs, and substantial attrition rates (Zhang et al., 2025). A central computational problem is therefore to estimate the binding affinity between candidate compounds and protein targets early enough to reduce the experimental search space (Schneider, 2018; Dara et al., 2022). Molecular docking can provide physically interpretable binding poses, but its computational cost and dependence on high-quality structural information limit its scalability for large candidate libraries. Learning-based drug-target affinity (DTA) prediction offers a complementary route for virtual screening, lead optimization, and drug repositioning (Yang et al., 2022a; Zeng et al., 2024).

A key challenge in DTA prediction is representation learning across heterogeneous biological and chemical information. Molecular encoders must preserve global graph topology while remaining sensitive to local functional patterns; protein representations must summarize long sequences without discarding informative residue context; and pair-specific or relational signals must retain information that isolated drug and protein descriptors cannot capture. Because these representations differ in source, scale, and semantics, directly summing or concatenating them may introduce redundant or conflicting information. The challenge becomes more pronounced when prediction involves previously unseen drugs or targets.

To address these issues, we propose PAFDTA, a heterogeneous representation-learning framework built around Progressive Adaptive Fusion and Cross-Branch Contrastive Alignment. PAFDTA learns complementary representations from molecular graphs, pretrained protein residue embeddings, joint drug-target token context, and heterogeneous relational topology. Multi-scale Local-Global Aggre-



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

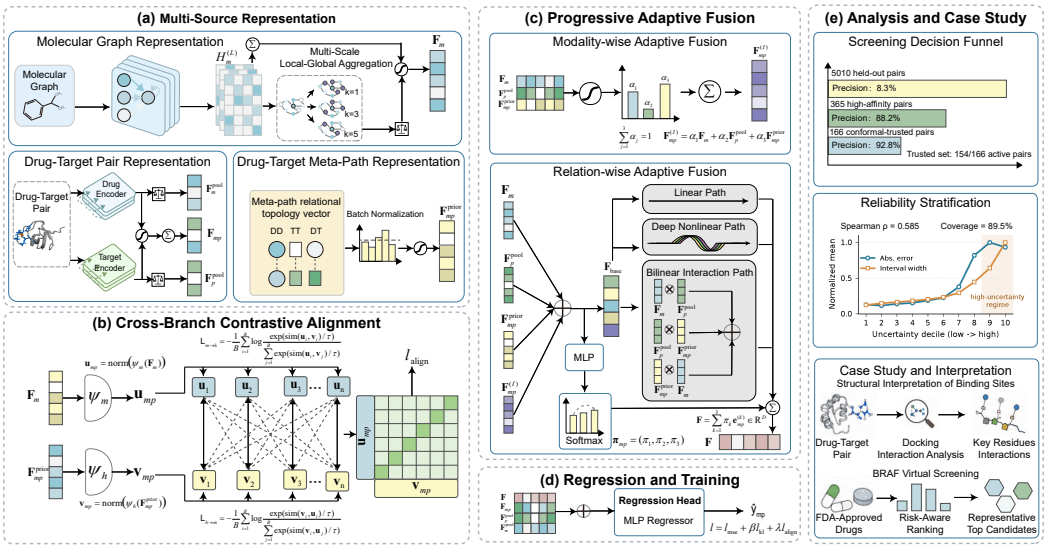



Figure 1: Overview of PAFDTA. The framework consists of (a) multi-source representation learning, (b) Cross-Branch Contrastive Alignment, (c) Progressive Adaptive Fusion, and (d) affinity regression. Uncertainty-aware screening and docking are used for downstream analysis.

gation (MLGA) enriches the molecular representation with global and local structural cues. Cross-Branch Contrastive Alignment (CBCA) uses pair-dependent relational context to regularize the drug-centric structural latent space before fusion. Progressive Adaptive Fusion then coordinates the factorized molecular, protein, and relational representations in two stages, while the joint drug-target representation is retained as a direct pair-context pathway to the regression head.

The main contributions are summarized as follows:

- We develop a multi-source representation-learning framework that jointly captures molecular structure, pretrained protein information, drug-target pair context, and high-order relational topology for DTA prediction.
- We introduce topology-guided contrastive regularization and Progressive Adaptive Fusion to coordinate heterogeneous representations before and during fusion, yielding a pair-conditioned affinity representation.
- We evaluate the learned representations and prediction framework through warm and cold-start benchmarks, component ablations, robustness and uncertainty analyses, and structural case studies on binding-site attention and BRAF-oriented repositioning.

## 2 Related Work

**Sequence- and graph-based DTA prediction.** Early DTA methods generally combined hand-crafted drug and protein descriptors with conventional machine-learning models. Representative examples include RF-Score based on random forests (Ballester & Mitchell, 2010), paired-input kernel learning (Pahikkala et al., 2015), and SimBoost, which integrates similarity features and matrix-factorization signals using gradient boosting (He et al., 2017). DeepDTA subsequently demonstrated that drug SMILES and protein sequences could be modeled end-to-end using convolutional neural networks (Öztürk et al., 2018). To better capture chemical topology, GraphDTA replaced the drug sequence encoder with graph neural networks (Nguyen et al., 2021), while MGraphDTA further explored multi-scale graph modeling to capture local and global molecular patterns (Yang et al., 2022b). More broadly, molecular representation learning has advanced through graph pre-training across node and graph levels (Hu et al., 2020), directional molecular message passing (Gasteiger et al., 2020), 2D–3D multi-view pre-training (Liu et al., 2022), context-aware molecular tokenization (Xia et al., 2023), and universal 3D pre-training (Zhou et al., 2023). These developments emphasize

representations that preserve complementary local chemistry and higher-order structural context. These studies motivate our graph-based molecular encoder with local-global aggregation.

**Interaction modeling and pretrained biological representations.** A second line of work focuses on richer drug-protein interaction modeling. AttentionDTA introduces attention mechanisms into sequence-based affinity prediction (Zhao et al., 2019); MolTrans models sub-structural interactions with Transformer-style representations (Huang et al., 2021); BACPI and PerceiverCPI further exploit attention-based cross-input dependencies (Li et al., 2022; Nguyen et al., 2023); and MRB-DTA combines multiple representation branches for DTA prediction (Zhang et al., 2022). More recently, pretrained protein language models have provided residue representations learned from large unlabeled sequence corpora. ESM2, in particular, offers transferable protein features that can be incorporated into downstream biological prediction tasks (Lin et al., 2023). LLMDTA further studies pretrained biological language-model representations in cold-start DTA settings (Tang et al., 2025). In PAFDTA, frozen ESM2 residue embeddings are combined with molecular atom tokens and relational context rather than being used as a standalone protein descriptor.

**Relational, contrastive, and multimodal learning.** Beyond isolated drug-target pairs, heterogeneous relations can provide high-order context. Affinity2Vec constructs a weighted heterogeneous drug-target graph and derives meta-path scores from drug-drug, target-target, and drug-target relations (Thafar et al., 2022). We follow these established relational schemas and integrate the resulting topology prior into an end-to-end representation-learning framework. Contrastive objectives are also widely used for graph representation learning, including mutual-information maximization in Deep Graph Infomax (Velickovic et al., 2019). In DTA, GraphCL-DTA learns molecular representations through semantics-preserving graph contrastive learning (Yang et al., 2024), while EMHGCL-DTA maximizes cross-scale mutual information in heterogeneous graphs (Kang et al., 2026). In parallel, multimodal fusion methods aim to coordinate complementary representations; for example, FMIN develops flexible iterative fusion with geometry-aware alignment (Zhang et al., 2026). Compared with these approaches, PAFDTA learns complementary structural, sequential, pairwise, and relational representations and coordinates them through topology-guided regularization and progressive fusion.

# 3 METHOD

Figure 1 presents the overall architecture of PAFDTA. For a drug-protein pair, the drug and protein are denoted by $m$ and $p$, respectively. The central objective is to learn complementary representations of molecular structure, protein sequence, pairwise interaction context, and relational topology, and then coordinate these representations before affinity regression.

## 3.1 MULTI-SOURCE REPRESENTATION LEARNING

### 3.1.1 MOLECULAR GRAPH REPRESENTATION

Each drug molecule $m$ is represented by a SMILES string and parsed by RDKit into an undirected molecular graph $G_m = (V_m, E_m)$, where atoms and chemical bonds are treated as nodes and edges. Let $X_m \in \mathbb{R}^{n_m \times F_a}$ denote the atom-feature matrix. A Graph Isomorphism Network (GIN) propagates atom and bond information through $L$ message-passing layers and produces the atom-level representation $H_m^{(L)}$.

To preserve complementary global and local molecular cues, MLGA summarizes the GIN representation through a graph-level branch and a multi-scale local branch:

$$\mathbf{F}_m^{\text{global}} = \text{SumPool}(H_m^{(L)}), \qquad \mathbf{F}_m^{\text{local}} = \text{AttnPool}(\text{MSConv}(H_m^{(L)})). \tag{1}$$

The local branch uses convolutional kernels of sizes 1, 3, and 5 over the deterministic atom ordering produced during preprocessing, whereas the GIN encoder propagates information through graph connectivity. A dimension-wise gate combines the two representations:

$$\mathbf{f}_m = \mathbf{q}_m \odot \mathbf{F}_m^{\text{local}} + (1 - \mathbf{q}_m) \odot \mathbf{F}_m^{\text{global}}, \tag{2}$$

where $\mathbf{q}_m = \sigma(\mathrm{MLP}_q([\mathbf{F}_m^{\mathrm{global}}; \mathbf{F}_m^{\mathrm{local}}]))$. A variational projection regularizes the molecular latent representation:

$$\mathbf{F}_m = \boldsymbol{\mu}_m + \boldsymbol{\sigma}_m \odot \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, I). \tag{3}$$

Here, $\boldsymbol{\mu}_m$ and $\log \boldsymbol{\sigma}_m^2$ are learned from $\mathbf{f}_m$, and $\mathbf{F}_m \in \mathbb{R}^D$ denotes the resulting molecular representation.

#### 3.1.2 Drug-Target Pair Representation

To construct joint drug-target token context, PAFDTA forms a pair representation from molecular atom tokens and protein residue tokens. For each target protein, residue-level features are precomputed using frozen ESM2-t30-150M-UR50D embeddings (640 dimensions) (Lin et al., 2023) and projected into the hidden dimension $D$. Long protein sequences are processed in windows during feature extraction, and the downstream model retains at most 1024 residue embeddings. The molecular atom tokens from the GIN encoder and projected protein residue tokens are separately contextualized by Transformer encoders with padding masks.

A masked attention-pooling operator summarizes the contextualized molecular and protein streams to obtain the pooled drug representation $\mathbf{F}_m^{\mathrm{pool}}$ and pooled protein representation $\mathbf{F}_p^{\mathrm{pool}}$. To preserve pair-level token context, the two contextualized token sequences are further concatenated and summarized by the same pooling mechanism, yielding the joint drug-target representation $\mathbf{F}_{mp} \in \mathbb{R}^D$. This branch preserves joint pair-level context beyond independent drug and protein pooling.

#### 3.1.3 Drug-Target Meta-Path Representation

To incorporate high-order relational context, PAFDTA follows the six heterogeneous path schemas DDT, DTT, DDDT, DTTT, DDTT, and DTDT used in Affinity2Vec (Thafar et al., 2022). Drug-drug, target-target, drug-target, and target-drug relations are first scaled and normalized to obtain transition matrices. The matrices are multiplied along each path schema to produce pair-level meta-path scores. For every pair $(m, p)$, the six scores are collected into a topology-prior vector and projected through batch normalization, a learnable linear transformation, and dropout to obtain $\mathbf{F}_{mp}^{\mathrm{prior}} \in \mathbb{R}^D$. For each experimental split, the drug-target relation is constructed only from training interactions, while drug-drug and target-target relations are derived from the available similarity information. The explicit path-matrix definitions are provided in Appendix C.

### 3.2 Cross-Branch Contrastive Alignment

The molecular representation $\mathbf{F}_m$ is drug-centric, whereas $\mathbf{F}_{mp}^{\mathrm{prior}}$ provides pair-dependent relational context. CBCA uses topology as an auxiliary signal rather than treating the branches as semantically identical. Two projection heads produce normalized vectors:

$$\mathbf{u}_i = \mathrm{norm}(\psi_m(\mathbf{F}_{m,i})), \qquad \mathbf{v}_i = \mathrm{norm}(\psi_h(\mathbf{F}_{mp,i}^{\mathrm{prior}})). \tag{4}$$

Each projection head is a two-layer MLP and $\mathrm{norm}(\mathbf{r}) = \mathbf{r}/\|\mathbf{r}\|_2$. Similarity and its temperature-scaled score are defined together as

$$\mathrm{sim}(\mathbf{x}_i, \mathbf{y}_j) = \frac{\mathbf{x}_i^\top \mathbf{y}_j}{\|\mathbf{x}_i\|_2 \|\mathbf{y}_j\|_2}, \qquad \eta_{ij} = \exp(\mathrm{sim}(\mathbf{x}_i, \mathbf{y}_j)/\tau). \tag{5}$$

For a mini-batch of size $B$, the directional InfoNCE objective is

$$\mathcal{L}_{x \to y} = -\frac{1}{B} \sum_{i=1}^{B} \log \frac{\eta_{ii}}{\sum_{j=1}^{B} \eta_{ij}}. \tag{6}$$

Matched structural-topology pairs serve as positives and mismatched samples in the mini-batch serve as negatives. The two directions are averaged as $\mathcal{L}_{\mathrm{align}} = \frac{1}{2}(\mathcal{L}_{m \to h} + \mathcal{L}_{h \to m})$. This regularizer encourages the molecular latent representation to remain compatible with the relational context associated with its observed drug-target pair before fusion.

### 3.3 PROGRESSIVE ADAPTIVE FUSION

Progressive Adaptive Fusion integrates the molecular representation $\mathbf{F}_m$, pooled protein representation $\mathbf{F}_p^{\text{pool}}$, and topology representation $\mathbf{F}_{mp}^{\text{prior}}$. Their branch-specific encoders produce dimension-matched representations in $\mathbb{R}^D$ that retain distinct branch semantics and are jointly optimized within the prediction model. Integration proceeds from source weighting to relation modeling.

**Modality-wise adaptive fusion.** The first stage summarizes each representation and maps the summary to normalized modality weights:

$$\mathbf{s}_{mp} = [\text{mean}(\mathbf{F}_m); \text{mean}(\mathbf{F}_p^{\text{pool}}); \text{mean}(\mathbf{F}_{mp}^{\text{prior}})], \qquad \boldsymbol{\alpha}_{mp} = \text{softmax}(\text{MLP}_\alpha(\mathbf{s}_{mp})). \tag{7}$$

With $\boldsymbol{\alpha}_{mp} = (\alpha_1, \alpha_2, \alpha_3)$ and $\sum_j \alpha_j = 1$, the modality representation is

$$\mathbf{F}_{mp}^{\text{mod}} = \alpha_1 \mathbf{F}_m + \alpha_2 \mathbf{F}_p^{\text{pool}} + \alpha_3 \mathbf{F}_{mp}^{\text{prior}}. \tag{8}$$

This stage replaces manually fixed source weights with learned pair-conditioned weighting over molecular, protein, and topology representations.

**Relation-wise adaptive fusion.** The second stage forms a joint base representation:

$$\mathbf{F}_{\text{base}} = [\mathbf{F}_m; \mathbf{F}_p^{\text{pool}}; \mathbf{F}_{mp}^{\text{prior}}; \mathbf{F}_{mp}^{\text{mod}}] \in \mathbb{R}^{4D}, \tag{9}$$

This base representation is processed by relation experts with different interaction complexity. The shallow and deep experts are

$$\mathbf{e}_{mp}^{(1)} = \text{MLP}_1(\mathbf{F}_{\text{base}}), \qquad \mathbf{e}_{mp}^{(2)} = \text{MLP}_2(\mathbf{F}_{\text{base}}), \tag{10}$$

while the multiplicative expert models explicit cross-source interactions:

$$\mathbf{e}_{mp}^{(3)} = \text{MLP}_3([\mathbf{F}_m \odot \mathbf{F}_p^{\text{pool}}; \mathbf{F}_m \odot \mathbf{F}_{mp}^{\text{prior}}; \mathbf{F}_p^{\text{pool}} \odot \mathbf{F}_{mp}^{\text{prior}}]). \tag{11}$$

A learned gate combines the expert representations to form the final affinity representation:

$$\boldsymbol{\pi}_{mp} = \text{softmax}(\text{MLP}_\pi(\mathbf{F}_{\text{base}})), \qquad \mathbf{F} = \sum_{k=1}^{3} \pi_k \mathbf{e}_{mp}^{(k)}. \tag{12}$$

Thus, the progressive design coordinates the factorized molecular, protein, and relational branches and refines their interactions into a task-specific fused representation. The joint pair representation $\mathbf{F}_{mp}$ is intentionally not compressed by these fusion stages; it is preserved as a direct pair-context pathway to the predictor.

### 3.4 REGRESSION AND TRAINING

The regression head combines the adaptive fusion pathway with the direct joint-pair pathway. It concatenates the pooled drug representation, joint pair representation, pooled protein representation, and final fused representation as $\mathbf{r}_{mp} = [\mathbf{F}_m^{\text{pool}}; \mathbf{F}_{mp}; \mathbf{F}_p^{\text{pool}}; \mathbf{F}] \in \mathbb{R}^{4D}$. A three-layer MLP, $4D \to 512 \to 128 \to 1$, then predicts $\hat{y}_{mp}$ using ReLU activations and dropout.

Training combines affinity regression with molecular latent regularization and cross-branch alignment. Mean squared error is $\mathcal{L}_{\text{mse}} = N^{-1} \sum_i (y_i - \hat{y}_i)^2$, while $\mathcal{L}_{\text{kl}}$ denotes the standard Gaussian KL regularizer for the variational molecular representation. The complete objective is

$$\mathcal{L} = \mathcal{L}_{\text{mse}} + \beta \mathcal{L}_{\text{kl}} + \lambda \mathcal{L}_{\text{align}}, \tag{13}$$

Here, $\beta$ and $\lambda$ weight the two auxiliary regularizers.

## 4 EXPERIMENTS

### 4.1 EXPERIMENTAL SETTINGS

We evaluate PAFDTA on the Davis (Davis et al., 2011) and KIBA (Tang et al., 2014) benchmarks. Davis contains 30,056 measured pairs from 68 drugs and 442 targets, while KIBA contains 118,254

Table 1: Comparison with representative DTA methods on KIBA and Davis. PAFDTA (Ours) reports the five-fold mean; baseline values are literature-reported. Best results are **bold** and second-best results are underlined. ↑: higher is better; ↓: lower is better.

| | KIBA | | | Davis | | |
|---|---|---|---|---|---|---|
| Method | MSE↓ | CI↑ | $R^2$ ↑ | MSE↓ | CI↑ | $R^2$ ↑ |
| GANsDTA (Zhao et al., 2020) | 0.2240 | 0.8660 | 0.6750 | 0.2760 | 0.8810 | 0.6530 |
| TF-DTA (Li et al., 2023) | 0.1771 | 0.8768 | 0.7344 | 0.2312 | 0.8856 | 0.6703 |
| GDilatedDTA (Zhang et al., 2024) | 0.1560 | 0.8760 | 0.7750 | 0.2370 | 0.8850 | 0.6860 |
| TransVAE-DTA (Zhou et al., 2024) | 0.2536 | 0.8221 | 0.6329 | 0.3329 | 0.8696 | 0.5713 |
| DTIAM (Lu et al., 2025) | 0.1620 | 0.8770 | – | 0.2120 | **0.9040** | – |
| LLMDTA (Tang et al., 2025) | 0.1620 | 0.8720 | 0.7680 | 0.2260 | 0.8840 | 0.7170 |
| PAFDTA (Ours) | **0.1500** | **0.8868** | **0.7858** | **0.2025** | 0.8994 | **0.7368** |

records from 2,111 drugs and 229 targets. The primary benchmark uses five-fold pair-wise cross-validation: the held-out folds are mutually exclusive and together cover every observed pair exactly once. Warm runs use seed 42, a fixed 200-epoch schedule, and final-epoch evaluation; Table 1 reports the five-fold mean. For each fold, relational features are constructed from its training interactions. We also evaluate drug-cold, target-cold, and double-cold splits on KIBA to examine representation transfer to unseen entities. Dataset processing and split construction are detailed in Appendix A.

The hidden dimension is set to $D = 192$, and the Transformer encoders use six attention heads. The optimizer is Adam with learning rate $5 \times 10^{-5}$ and weight decay $10^{-5}$, with dropout set to 0.25. The maximum number of training epochs is 200. Protein residue features are precomputed using frozen ESM2-t30-150M-UR50D embeddings; further implementation details are given in Appendix B.

### 4.2 Evaluation Metrics

We use mean squared error (MSE), concordance index (CI), and the coefficient of determination ($R^2$), which are widely adopted for DTA regression. MSE measures numerical affinity error, CI evaluates whether predicted affinities preserve the relative ordering of sample pairs, and $R^2$ measures goodness of fit. Lower MSE and higher CI/$R^2$ indicate better performance. For the uncertainty-aware screening analysis, we additionally report candidate retention, precision, recall, and empirical interval coverage.

### 4.3 Benchmark Comparison

We compare PAFDTA with representative DTA methods covering adversarial learning, Transformer-based sequence modeling, graph neural networks, variational representation learning, attention-based interaction modeling, and pretrained biological language models. The compared methods include GANsDTA (Zhao et al., 2020), TF-DTA (Li et al., 2023), GDilatedDTA (Zhang et al., 2024), TransVAE-DTA (Zhou et al., 2024), DTIAM (Lu et al., 2025), and LLMDTA (Tang et al., 2025). Table 1 reports the benchmark results on KIBA and Davis. Values for comparison methods are taken from the corresponding publications, following the reporting convention of prior DTA studies.

Averaged over the five warm folds, PAFDTA obtains 0.1500 MSE, 0.8868 CI, and 0.7858 $R^2$ on KIBA. On Davis, it reaches 0.2025 MSE, 0.8994 CI, and 0.7368 $R^2$. Relative to the literature-reported methods in Table 1, PAFDTA compares favorably, achieving the lowest listed MSE on both datasets while remaining competitive in concordance ranking. These results show that coordinating complementary molecular, protein, pairwise, and relational representations provides strong warm-start affinity prediction across both benchmarks.

### 4.4 Cold-Start Evaluation

We further evaluate PAFDTA under three entity-level KIBA settings: drug-cold, where test drugs are unseen during training; target-cold, where test targets are unseen; and double-cold, where both sides are held out. All methods are evaluated on the same fixed split for each setting with the same target

Table 2: Cold-start comparison on KIBA with LLMDTA (Tang et al., 2025), TransVAE-DTA (Zhou et al., 2024), and DTIAM (Lu et al., 2025). Best results are **bold** and second-best results are underlined.

| | Drug-cold | | | Target-cold | | | Double-cold | | |
|---|---|---|---|---|---|---|---|---|---|
| Method | MSE↓ | CI↑ | $R^2$ ↑ | MSE↓ | CI↑ | $R^2$ ↑ | MSE↓ | CI↑ | $R^2$ ↑ |
| TransVAE-DTA (Zhou et al., 2024) | 0.548 | 0.679 | 0.281 | 0.535 | 0.660 | 0.117 | 0.714 | 0.629 | -0.065 |
| LLMDTA (Tang et al., 2025) | 0.387 | 0.770 | 0.492 | **0.312** | 0.747 | **0.485** | **0.482** | 0.681 | **0.281** |
| DTIAM (Lu et al., 2025) | 0.436 | 0.765 | 0.428 | 0.389 | 0.689 | 0.358 | 0.566 | 0.617 | 0.155 |
| PAFDTA (Ours) | **0.379** | **0.791** | **0.503** | 0.319 | **0.751** | 0.474 | 0.543 | **0.687** | 0.189 |

Table 3: Ablation study on Davis. Best results are shown in **bold**, and second-best results are underlined.

| Ablation type | Variant | MSE↓ | CI↑ | $R^2$ ↑ |
|---|---|---|---|---|
| Modality ablation | w/o $\mathbf{F}_m$ | 0.2061 | 0.8936 | 0.7343 |
| | w/o $\mathbf{F}_{mp}$ | 0.2079 | 0.9002 | 0.7320 |
| | w/o $\mathbf{F}_{mp}^{\text{prior}}$ | 0.2051 | **0.9047** | 0.7355 |
| Module ablation | w/o CBCA | 0.2045 | 0.9043 | 0.7363 |
| | w/o MLGA | 0.2093 | 0.8999 | 0.7302 |
| Fusion method | Sum | 0.2084 | 0.8980 | 0.7313 |
| | Concat | 0.2064 | 0.8998 | 0.7339 |
| Full model | PAFDTA (Ours) | **0.2018** | 0.8980 | **0.7376** |

values and metric implementation. Table 2 compares PAFDTA with LLMDTA (Tang et al., 2025), TransVAE-DTA (Zhou et al., 2024), and DTIAM (Lu et al., 2025). Additional split and baseline details are provided in Appendix D.

Under drug-cold evaluation, PAFDTA obtains the best result on all three metrics, reaching 0.791 CI with an MSE of 0.379 and an $R^2$ of 0.503. Under target-cold evaluation, PAFDTA achieves the highest CI of 0.751. A similar pattern is observed under the more challenging double-cold setting: PAFDTA achieves the highest CI of 0.687. These results indicate that the learned multi-source representations retain competitive affinity ranking when drugs and/or targets are unseen, complementing the warm benchmark evaluation.

### 4.5 Ablation Study

We examine the contribution of the principal representation and fusion components on Davis. The variants remove individual representation or regularization components and replace Progressive Adaptive Fusion with fixed fusion rules. Table 3 reports MSE, CI, and $R^2$ for the resulting variants.

The full model achieves the lowest MSE (0.2018) and highest $R^2$ (0.7376), indicating the strongest overall numerical affinity fit. Removing the relational prior or CBCA yields slightly higher CI (0.9047 and 0.9043, respectively), showing that concordance ranking and point-wise regression error respond differently to the representation components. Replacing Progressive Adaptive Fusion with fixed Sum or Concat fusion increases MSE and reduces $R^2$ relative to the full model, supporting the value of learned fusion for numerical affinity prediction. Additional robustness checks on molecular node reindexing and cold-start structural fingerprint augmentation are provided in Appendix E.

### 4.6 Prediction Consistency and Uncertainty-Aware Screening

We further analyze prediction consistency on Davis. The residual distribution and predicted-versus-true affinity plot are provided in Appendix F. Most predictions are concentrated near the reference line: 73.5%, 83.9%, and 94.4% of samples have absolute errors below 0.25, 0.50, and 1.00 pKd, respectively. The median absolute error is 0.067 pKd, with 90th/95th percentiles of 0.734/1.070 pKd.

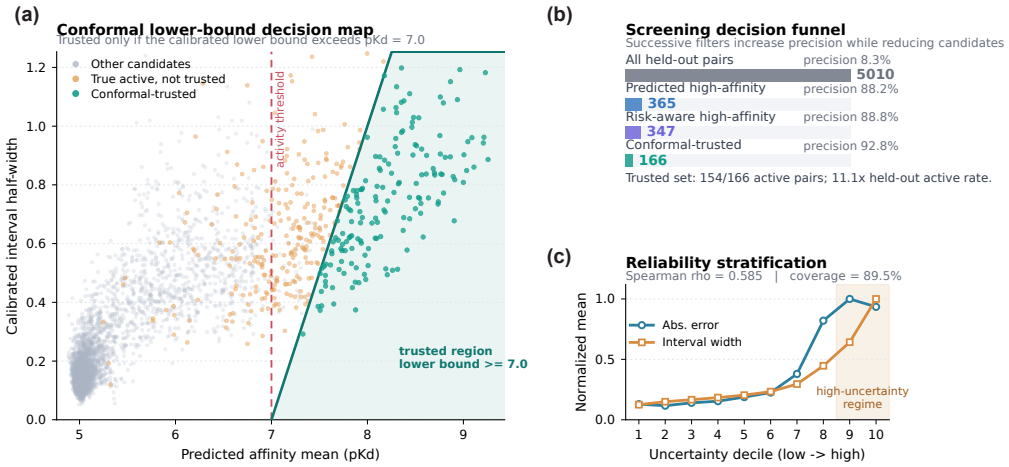



Figure 2: Uncertainty-aware virtual-screening analysis on Davis. (a) Conformal lower-bound decision map. (b) Candidate-screening funnel. (c) Reliability trends across uncertainty deciles.

To examine practical candidate prioritization, we additionally conduct an uncertainty-aware virtual-screening analysis on Davis. Drug-target pairs with pKd$\geq 7.0$ are regarded as active. Starting from 5,010 held-out pairs, 365 pass the predicted-affinity threshold, 347 remain after uncertainty filtering, and 166 satisfy the conformal lower-bound criterion. Among the final candidates, 154 are true active pairs, corresponding to a precision of 92.8%. Relative to the 417 active pairs in the held-out set, the final stage retains 36.9% of the actives. The empirical interval coverage is 89.5% for the nominal 90% target.

Figure 2 summarizes the screening behavior from three complementary perspectives. The conformal lower-bound map shows how predictive uncertainty changes the acceptance boundary near the activity threshold, while the funnel makes the precision–retention trade-off explicit across successive filtering stages. The reliability analysis further shows that prediction errors and calibrated interval widths increase toward higher-uncertainty deciles. Together, these results support uncertainty as a useful representation-level reliability signal for candidate prioritization. The construction of the stochastic uncertainty score and the disjoint conformal calibration protocol are detailed in Appendix G.

### 4.7 Case Study

**Structural interpretation of learned protein representations.** To examine whether the learned target representation emphasizes structurally relevant regions, we compare high-attention residues with ligand-binding pocket residues derived from four co-crystal structures. Pocket residues are defined by a 5 Å heavy-atom distance to the bound ligand. Figure 3(a) shows that several highlighted residues coincide with or lie near known pocket regions, including PHE-143, VAL-269, and ASP-270 for 4Y85; CYS-39 for 3GC8; HIS-307 for 6EG9; and PHE-512, PRO-484, and ILE-521 for 8SSN. The correspondence provides qualitative structural evidence that the learned protein representation can concentrate on binding-relevant sequence regions rather than distributing attention uniformly.

**BRAF-oriented drug repositioning.** We further select BRAF (PDB ID: 4RZV), a kinase associated with tumor proliferation, and rank a library of 2,138 FDA-approved small molecules (Pihan et al., 2012). High-ranked candidates are subsequently evaluated with AutoDock Vina (Ding et al., 2023). Zongertinib, M Midostaurin, Avapritinib, and Midostaurin rank among the leading candidates and obtain Vina energies of $-8.9$, $-8.7$, $-8.3$, and $-8.3$ kcal/mol, respectively. As illustrated in Fig. 3(b), the four compounds occupy a similar BRAF cavity and show recurring contacts around residues such as HIS-509/HIS-510, VAL-511, ASP-587, THR-589, and LYS-591. These docking

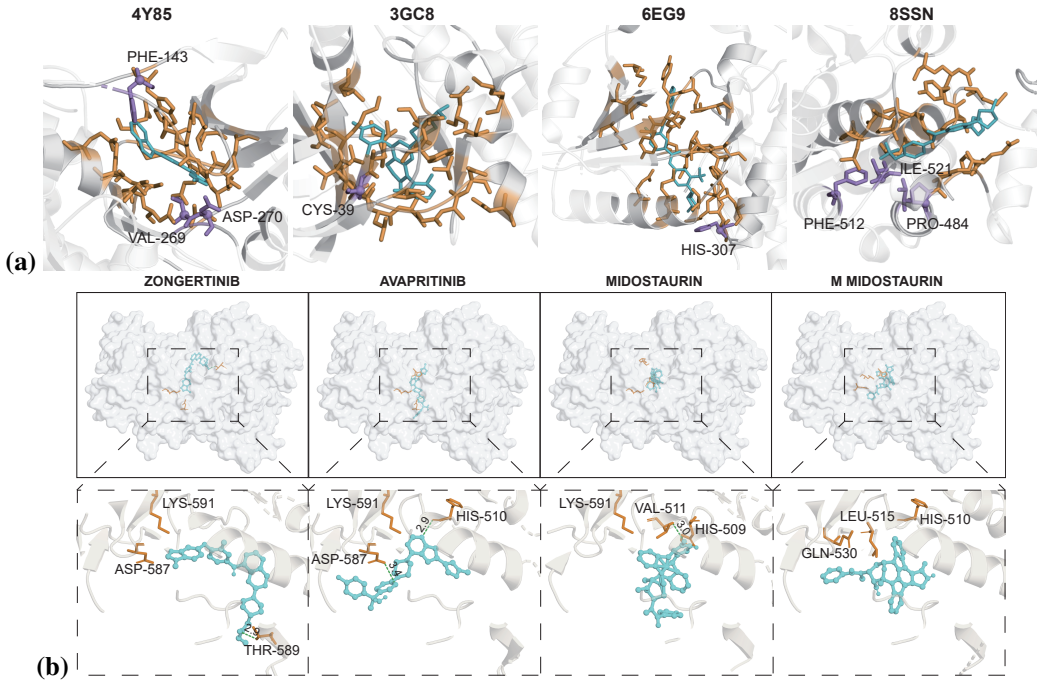



Figure 3: Case-study visualization. (a) Protein-attention and pocket-proximity examples for four co-crystal structures. (b) Docking poses of four high-ranked BRAF candidates, shown at global and local pocket scales.

poses complement the learned affinity ranking with structural hypotheses rather than experimental evidence.

## 5 Limitations and Future Work

Several directions can further strengthen the representation-learning framework. Future work will evaluate broader distribution shifts, including scaffold-aware drug splits, target-family shifts, and additional random seeds. The current relational branch relies on predefined drug-drug, target-target, and drug-target relations; learning task-adaptive relations or incorporating richer structural and biological knowledge may improve relational representations. Extending the molecular and protein encoders with three-dimensional conformations and structure-aware protein features is another promising direction. Finally, prospective assays would provide stronger validation for candidates prioritized by uncertainty-aware screening and docking.

## 6 Conclusion

We proposed PAFDTA, a heterogeneous representation-learning framework for DTA prediction. It learns complementary molecular, protein, pairwise, and relational representations, uses topology-guided contrastive regularization to coordinate structural and relational context, and integrates the learned representations through Progressive Adaptive Fusion. Experiments on Davis and KIBA demonstrate strong warm and cold-start affinity prediction, while component ablations, uncertainty analysis, and structural case studies provide complementary evidence for the learned representation framework. Future work will explore broader distribution shifts, richer structural representations, and prospective validation of model-prioritized candidates.

### AI USE STATEMENT

Generative AI tools were used for manuscript organization, language editing, and consistency checking between the manuscript and implementation. The authors reviewed and independently verified all AI-assisted technical changes against the underlying code, data, experimental records, and cited literature. Generative AI was not treated as a source of experimental measurements or empirical results. The authors take responsibility for the final content, claims, and artifacts in this submission.

### ETHICS STATEMENT

This work uses public drug-target benchmark data and computational analyses and does not involve human subjects or clinical intervention. The virtual-screening and docking results are hypothesis-generating computational outputs and should not be interpreted as medical advice or clinical evidence.

### REPRODUCIBILITY STATEMENT

The main paper describes the model architecture, training objective, datasets, evaluation metrics, and primary experiments. The appendix provides hyperparameters, evaluation protocols, relational-prior construction, cold-start settings, robustness analyses, prediction-consistency results, uncertainty-aware screening details, and the complete docking study.

## REFERENCES

Pedro J Ballester and John BO Mitchell. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics*, 26(9):1169–1175, 2010.

Suresh Dara, Swetha Dhamercherla, Surender Singh Jadav, CH Madhu Babu, and Mohamed Jawed Ahsan. Machine learning in drug discovery: a review. *Artificial intelligence review*, 55(3):1947–1999, 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.

Ji Ding, Shidi Tang, Zheming Mei, Lingyue Wang, Qinqin Huang, Haifeng Hu, Ming Ling, and Jiansheng Wu. Vina-gpu 2.0: further accelerating autodock vina and its derivatives with graphics processing units. *Journal of chemical information and modeling*, 63(7):1982–1998, 2023.

Johannes Gasteiger, Janek Gross, and Stephan Guennemann. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020.

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.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. In *International Conference on Learning Representations*, 2020.

Kexin Huang, Cao Xiao, Lucas M Glass, and Jimeng Sun. Moltrans: molecular interaction transformer for drug–target interaction prediction. *Bioinformatics*, 37(6):830–836, 2021.

Xiao Kang, Licai Zhang, Zekun Wang, Xinxing Yang, Lin Wang, Genke Yang, and Jian Chu. Enhancing cross-scale feature mutual information via heterogeneous graph contrastive learning for drug-target binding affinity prediction. *IEEE Journal of Biomedical and Health Informatics*, 2026. doi: 10.1109/JBHI.2026.3702061. Online ahead of print.

Min Li, Zhangli Lu, Yifan Wu, and YaoHang Li. Bacpi: a bi-directional attention neural network for compound–protein interaction and binding affinity prediction. *Bioinformatics*, 38(7):1995–2002, 2022.

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)*, pp. 418–421. IEEE, 2023.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

Shengchao Liu, Hanchen Wang, Weiyang Liu, Joan Lasenby, Hongyu Guo, and Jian Tang. Pre-training molecular graph representation with 3d geometry. In *International Conference on Learning Representations*, 2022.

Zhangli Lu, Guoqiang Song, Huimin Zhu, Chuqi Lei, Xinliang Sun, Kaili Wang, Libo Qin, Yafei Chen, Jing Tang, and Min Li. Dtiam: a unified framework for predicting drug-target interactions, binding affinities and drug mechanisms. *Nature Communications*, 16(1):2548, 2025.

Ngoc-Quang Nguyen, Gwanghoon Jang, Hajung Kim, and Jaewoo Kang. Perceiver cpi: a nested cross-attention network for compound–protein interaction prediction. *Bioinformatics*, 39(1): btac731, 2023.

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.

Emilie Pihan, Lionel Colliandre, Jean-François Guichou, and Dominique Douguet. e-drug3d: 3d structure collections dedicated to drug repurposing and fragment-based drug design. *Bioinformatics*, 28(11):1540–1541, 2012.

Gisbert Schneider. Automating drug discovery. *Nature reviews drug discovery*, 17(2):97–113, 2018.

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*, 22(6):2398–2409, 2025. doi: 10.1109/TCBBIO.2025.3541634.

Maha A. Thafar, Mona Alshahrani, Somayah Albaradei, Takashi Gojobori, Magbubah Essack, and Xin Gao. Affinity2vec: drug-target binding affinity prediction through representation learning, graph mining, and machine learning. *Scientific Reports*, 12:4751, 2022. doi: 10.1038/s41598-022-08787-9.

Petar Velickovic, William Fedus, William L. Hamilton, Pietro Lio, Yoshua Bengio, and R. Devon Hjelm. Deep graph infomax. In *International Conference on Learning Representations*, 2019.

Jun Xia, Chengshuai Zhao, Bozhen Hu, Zhangyang Gao, Cheng Tan, Yue Liu, Siyuan Li, and Stan Z. Li. Mole-bert: Rethinking pre-training graph neural networks for molecules. In *International Conference on Learning Representations*, 2023.

Xinxing Yang, Genke Yang, and Jian Chu. Graphcl-dta: A graph contrastive learning with molecular semantics for drug-target binding affinity prediction. *IEEE Journal of Biomedical and Health Informatics*, 28(8):4544–4552, 2024. doi: 10.1109/JBHI.2024.3350666.

Xixi Yang, Zhangming Niu, Yuansheng Liu, Bosheng Song, Weiqiang Lu, Li Zeng, and Xiangxiang Zeng. Modality-dta: multimodality fusion strategy for drug–target affinity prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 20(2):1200–1210, 2022a.

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, 2022b.

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.

Kang Zhang, Xin Yang, Yifei Wang, Yunfang Yu, Niu Huang, Gen Li, Xiaokun Li, Joseph C Wu, and Shengyong Yang. Artificial intelligence in drug development. *Nature medicine*, 31(1):45–59, 2025.

Li Zhang, Chun-Chun Wang, and Xing Chen. Predicting drug–target binding affinity through molecule representation block based on multi-head attention and skip connection. *Briefings in Bioinformatics*, 23(6):bbac468, 2022.

Licai Zhang, Xiao Kang, Xinxing Yang, Lin Wang, Genke Yang, and Jian Chu. Fmin: A flexible multimodal iterative fusion network with geometry-aware positive noise alignment for drug-target interaction prediction. *Neurocomputing*, 681:133351, 2026. doi: 10.1016/j.neucom.2026.133351.

Longxin Zhang, Wenliang Zeng, Jingsheng Chen, Jianguo Chen, and Keqin Li. Gdilateddta: Graph dilation convolution strategy for drug target binding affinity prediction. *Biomedical Signal Processing and Control*, 92:106110, 2024.

Lingling Zhao, Junjie Wang, Long Pang, Yang Liu, and Jun Zhang. Gansdta: Predicting drug-target binding affinity using gans. *Frontiers in genetics*, 10:1243, 2020.

Qichang Zhao, Fen Xiao, Mengyun Yang, Yaohang Li, and Jianxin Wang. Attentiondta: prediction of drug–target binding affinity using attention model. In *2019 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pp. 64–69. IEEE, 2019.

Changjian Zhou, Zhongzheng Li, Jia Song, and Wensheng Xiang. Transvae-dta: Transformer and variational autoencoder network for drug-target binding affinity prediction. *Computer Methods and Programs in Biomedicine*, 244:108003, 2024.

Gengmo Zhou, Zhifeng Gao, Qiankun Ding, Hang Zheng, Hongteng Xu, Zhewei Wei, Linfeng Zhang, and Guolin Ke. Uni-mol: A universal 3d molecular representation learning framework. In *International Conference on Learning Representations*, 2023.

Table 4: Main implementation hyperparameters.

| Hyperparameter | Setting |
|---|---|
| Hidden dimension $D$ | 192 |
| Transformer attention heads | 6 |
| Dropout | 0.25 |
| Optimizer | Adam |
| Learning rate | $5 \times 10^{-5}$ |
| Weight decay | $10^{-5}$ |
| Maximum epochs | 200 |
| KL coefficient $\beta$ | 0.01 |
| Contrastive coefficient $\lambda$ | 0.05 |
| Davis batch size | 16 |
| KIBA batch size | 64 |
| Protein encoder | ESM2-t30-150M-UR50D |
| Protein feature dimension | 640 |
| Maximum residues in DTA model | 1024 |

# A Dataset and Evaluation Protocols

**Datasets.** Davis contains 68 drugs and 442 kinase targets with 30,056 measured interactions. The original dissociation constants are transformed to the pKd scale for regression. KIBA contains 2,111 drugs and 229 targets with 118,254 observed affinity records. Both datasets are widely used for benchmarking DTA models and provide complementary characteristics: Davis is relatively small and dense, whereas KIBA contains a larger set of compounds and sparse observed interactions.

**Warm evaluation.** The primary benchmark experiment follows five-fold pair-wise cross-validation generated with fixed seed 42. The five held-out folds are mutually exclusive, their union equals all observed drug-target pairs (30,056 for Davis and 118,254 for KIBA), and each observed pair appears in exactly one held-out fold. In each fold, the remaining four folds provide the training data. All runs use a fixed 200-epoch schedule, and the reported warm results use the final-epoch predictions rather than selecting a checkpoint by held-out performance. Drug-target topology priors are reconstructed from the corresponding training interactions. We report the mean across the five folds for MSE, CI, and $R^2$.

**Cold-start evaluation.** For the additional KIBA experiments, the data are split at the entity level. In drug-cold evaluation, test drugs do not occur in the training data; in target-cold evaluation, test targets are unseen during training; and in double-cold evaluation, both sides are held out. The split generator constructs complementary entity-level partitions with a fixed seed; the reported comparison uses the same fixed partition for every method within each setting. Relational priors are reconstructed only from the corresponding training interactions. All methods use the same target values and metrics.

**Metrics.** For $N$ samples, MSE is $\frac{1}{N}\sum_i (y_i - \hat{y}_i)^2$. CI measures the probability that the ordering of two predictions agrees with the ordering of their observed affinities. The coefficient of determination is computed as $R^2 = 1 - \sum_i (y_i - \hat{y}_i)^2 / \sum_i (y_i - \bar{y})^2$. For screening, precision is the active fraction among selected candidates, and recall is the retained fraction of all active pairs.

# B Implementation Details

Table 4 summarizes the main training configuration. Protein features are precomputed with frozen ESM2-t30-150M-UR50D. The feature-extraction stage uses windows for sequences beyond the ESM2 token limit, while the DTA model receives at most 1024 residue embeddings per sample. The hidden representation dimension is 192 throughout the main fusion pathway.

Table 5: Node-reindexing robustness diagnostic on 300 samples per setting. ΔCI is computed after a connectivity-preserving random node permutation.

| Setting | Mean $\lvert\Delta\hat{y}\rvert$ | P95 $\lvert\Delta\hat{y}\rvert$ | Corr. | ΔCI |
|---|---|---|---|---|
| KIBA drug-cold | $4.40 \times 10^{-4}$ | $1.83 \times 10^{-3}$ | 0.999998 | $-2.48 \times 10^{-4}$ |
| KIBA double-cold | $1.77 \times 10^{-3}$ | $7.31 \times 10^{-3}$ | 0.999985 | $-4.69 \times 10^{-4}$ |
| Davis drug-cold | $1.34 \times 10^{-5}$ | $7.30 \times 10^{-5}$ | $> 0.999999$ | 0 |

## C Relational Prior Construction

The relational branch starts from drug-drug similarity, target-target similarity, and observed drug-target affinity relations. After scaling and normalization, these relations define transition matrices $P_{DD}$, $P_{TT}$, $P_{DT}$, and $P_{TD}$. We follow the six path schemas DDT, DTT, DDDT, DTTT, DDTT, and DTDT used in Affinity2Vec (Thafar et al., 2022). The two shortest path-score matrices are

$$S_{DDT} = P_{DD}P_{DT}, \qquad S_{DTT} = P_{DT}P_{TT}. \tag{14}$$

Longer homogeneous extensions are obtained by adding one drug-drug or target-target transition:

$$S_{DDDT} = P_{DD}^2 P_{DT}, \qquad S_{DTTT} = P_{DT}P_{TT}^2. \tag{15}$$

The remaining mixed paths are computed as

$$S_{DDTT} = P_{DD}P_{DT}P_{TT}, \qquad S_{DTDT} = P_{DT}P_{TD}P_{DT}. \tag{16}$$

For each fold, $P_{DT}$ is built only from training interactions before the corresponding path matrices are multiplied. Pair-level path scores are concatenated, normalized, and projected to dimension $D$ before being passed to CBCA and Progressive Adaptive Fusion. This construction regenerates the relational representation for each split while leaving the downstream architecture unchanged.

## D Cold-Start Baseline Details

The KIBA cold-start comparison uses the same entity-level partition for all methods within each setting. For PAFDTA, the cold-start experiments use the same overall architecture and training objective as the main model. The test sets contain 22,374 drug-cold pairs, 24,418 target-cold pairs, and 4,637 double-cold pairs.

## E Node-Order Robustness

The local MLGA convolution operates on the deterministic atom ordering produced during molecular preprocessing, so we additionally test sensitivity to node reindexing. For each diagnostic setting, 300 molecules are randomly permuted while applying the identical permutation to node features and edge endpoints, preserving graph connectivity and bond attributes. Table 5 shows that the resulting prediction changes are very small: the 95th-percentile absolute change remains below $7.4 \times 10^{-3}$ pKd, prediction correlations exceed 0.99998, and the absolute CI change is below $5 \times 10^{-4}$ across the tested settings. This empirical diagnostic does not claim that the convolution itself is mathematically permutation invariant; rather, it indicates limited sensitivity of the trained predictor to atom reindexing in these evaluations.

## F Prediction Consistency Analysis

Figure 4 provides two complementary views of the Davis prediction errors. The cumulative residual distribution shows that 73.5%, 83.9%, and 94.4% of samples have absolute errors below 0.25, 0.50, and 1.00 pKd, respectively. The median absolute error is 0.067 pKd, and the 90% and 95% quantiles are 0.734 and 1.070 pKd. The predicted-versus-true scatter plot shows that the majority of points follow the $y = x$ trend, with larger deviations concentrated in a smaller set of difficult samples.

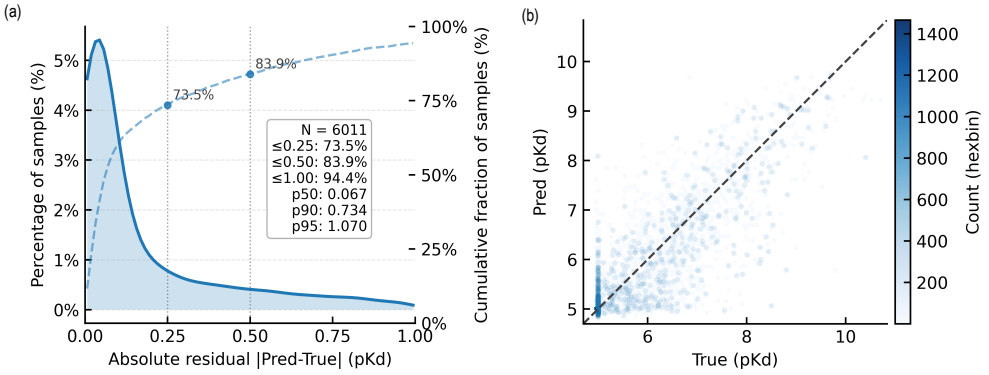



Figure 4: Prediction consistency on Davis. Left: absolute residual distribution and cumulative fraction. Right: predicted versus true pKd values with the $y = x$ reference line.

## G Uncertainty-Aware Screening Details

The uncertainty analysis is applied after affinity prediction and is not an additional trainable branch of PAFDTA. For each pair, we perform $T = 20$ stochastic forward passes with dropout active and the variational drug encoder sampled. Let $\hat{y}_i^{(t)}$ denote the $t$-th prediction. Their predictive mean is

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} \hat{y}_i^{(t)}. \tag{17}$$

The sample-specific uncertainty scale is the standard deviation across these stochastic predictions:

$$u_i = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} \left( \hat{y}_i^{(t)} - \bar{y}_i \right)^2}. \tag{18}$$

A disjoint Davis validation set of 5,010 pairs is used only for conformal calibration, while a separate 5,010-pair test set is used for screening evaluation. On the calibration set, the normalized residual is

$$r_i = \frac{|y_i - \bar{y}_i|}{u_i + \epsilon}, \qquad \epsilon = 10^{-8}. \tag{19}$$

For the 90% split-conformal interval, we use the finite-sample rank $k = \lceil (n_{\text{cal}} + 1)(1 - 0.1) \rceil = 4510$, which gives $q_{0.90} = 2.7241$. The calibrated interval for a test sample is

$$[\bar{y}_i - q_{0.90} u_i,\ \bar{y}_i + q_{0.90} u_i]. \tag{20}$$

A candidate is considered conformal-trusted when the lower bound of this interval remains above the activity threshold pKd$= 7.0$.

Starting from the 5,010 test pairs, 365 satisfy the predicted-affinity threshold and 347 remain after uncertainty filtering. The conformal lower-bound rule retains 166 candidates, of which 154 are active, yielding 92.8% precision. The test set contains 417 active pairs, so the final trusted set corresponds to 36.9% recall. The empirical test-set coverage is 89.5% for the nominal 90% target. The main paper visualizes the corresponding decision map, screening funnel, and uncertainty-decile reliability trend; this appendix records the calibration procedure and numerical definitions needed to reproduce that analysis.

## H BRAF Candidate Ranking Details

The main paper presents representative protein-pocket correspondence and BRAF docking poses. This supplement reports the complete top-10 BRAF candidate list used in the case study. After preprocessing, 2,138 FDA-approved small molecules (Pihan et al., 2012) are ranked by the trained model, and selected high-ranking compounds are evaluated with AutoDock Vina (Ding et al., 2023).

Table 6: Top 10 predicted BRAF candidate drugs. The model score is used for ranking; Vina energies are reported in kcal/mol.

| Rank | Drug | Model score | Vina score |
|---|---|---|---|
| 1 | Zongertinib | 2.22 | -8.9 |
| 2 | M Midostaurin | 2.05 | -8.7 |
| 3 | Avapritinib | 2.01 | -8.3 |
| 4 | Midostaurin | 1.92 | -8.3 |
| 5 | Tivozanib | 1.50 | -8.2 |
| 6 | Raloxifene | 1.50 | -8.2 |
| 7 | Regorafenib | 1.46 | -8.2 |
| 8 | Imatinib | 1.31 | -8.1 |
| 9 | Abemaciclib | 1.27 | -8.0 |
| 10 | Sorafenib | 1.17 | -7.8 |

The top-ranked set includes several kinase inhibitors and anticancer compounds. The representative candidates visualized in the main paper occupy a similar BRAF binding cavity and exhibit recurring local contacts around HIS-509/HIS-510, VAL-511, ASP-587, THR-589, and LYS-591. The docking scores and poses are used only as structural hypotheses supporting the ranking analysis, not as experimental validation.