Supplementary Information

MolRes-DTA: a molecular-multiview fusion and residue-aware model for drug–target affinity prediction

Hongli Hou,[a,b] Qi Wei,[b] Dian Huang,[*b] Minglu Zhao,[b] Hongliang Duan,[a] and Shengzhong Feng[*b]

[a] Faculty of Applied Sciences, Macao Polytechnic University, R. de Luís Gonzaga Gomes, 999078, Macao, China.

[b] Guangdong Institute of Intelligence Science and Technology, Hengqin, Zhuhai, 519031, Guangdong, China.

*corresponding to: huangdian@gdiist.cn; fengshengzhong@gdiist.cn

A. Dataset

(1) Dataset partitioning

Dataset partitioning details are provided in Table S1. For fair comparison and reproducibility, we used the fixed split protocol released by TEFDTA for all datasets. For Davis and KIBA, TEFDTA randomly shuffles and splits drug-target interaction pairs into training and test sets with an approximately 5:1 ratio (interaction-level split), ensuring that the same interaction record/pair does not appear in both sets. For BindingDB, we used the predefined TEFDTA benchmark split without re-splitting to maintain strict comparability. Since TEFDTA does not specify whether this predefined split is random, cluster-based, or time-based, we report it as a benchmark-provided predefined partition.

Table S1 Statistical analysis of benchmark datasets and the division of training and test sets

Dataset	No. of compounds	No. of proteins	No. of interactions	Training set	Test set
KIBA	2111	229	118254	98,545	19709
Davis	68	442	30056	25,046	5010
BindingDB	80324	5561	1254402	1234402	20000

We also used an alternative splitting approach on the Davis dataset to compare our model with two baselines, with results in Table S2 demonstrating that MolRes-DTA consistently outperforms the two baselines across key predictive metrics, achieving lower MSE and more reliable performance in DTA prediction.

Table S2 Results on the scaffold-based split (Davis)

Dataset	MSE (↓)	CI (↑)	r_m^2 (↑)
GraphDTA	0.193	0.889	-
TEFDTA	0.190	0.896	0.756
MolRes-DTA (ours)	**0.159**	**0.906**	**0.793**

(2) Discussion on potential biases of the source datasets

(i) Davis exhibits a highly skewed affinity distribution, with many weak/non-binding interactions concentrated around pKd≈5, which can bias learning toward weak binders and reduce predictive accuracy for strong-binding samples;
(ii) TEFDTA reported that earlier Davis releases used by some baselines contained cases where mutated protein sequences were mislabeled as wild-type, potentially introducing systematic bias due to incorrect sequence annotations (here we use the TEFDTA-corrected Davis dataset);
(iii) for KIBA, the original dataset (467 targets and 52,498 drugs) was filtered in prior work (DeepDTA) to retain only drugs and targets with at least 10 measured interactions, yielding 229 proteins and 2,111 drugs, which can underrepresent low-frequency compounds/targets and bias evaluation toward well-studied entities with richer interaction profiles. Other than adopting TEFDTA's corrected Davis sequences, we did not apply additional dataset-level bias mitigation strategies.

B. Significance analysis

We conducted paired t-tests to assess statistical significance across random seeds. The results confirm that MolRes-DTA achieves statistically significant improvements over both GraphDTA and TEFDTA ($p < 0.01$) on Davis and KIBA datasets, demonstrating that the observed gains are robust rather than random fluctuations.

Table S3 Significant comparison with the baseline model

Dataset	Comparison	t-statistic	p-value	Significance
Davis	GraphDTA vs MolRes-DTA	36.73	3.28×10^{-6}	$p < 0.001$
Davis	TEFDTA vs MolRes-DTA	17.12	6.83×10^{-5}	$p < 0.001$
KIBA	GraphDTA vs MolRes-DTA	44.99	1.46×10^{-6}	$p < 0.001$
KIBA	TEFDTA vs MolRes-DTA	39.81	2.38×10^{-6}	$p < 0.001$

C. Performance comparison of drug fingerprints

To evaluate the effect of standard molecular fingerprints, we compared MACCS and Morgan (ECFP4) fingerprints as the drug representation on the Davis dataset.

Table S4 Performance comparison of drug fingerprints on the Davis dataset.

Fingerprint Category	MSE (↓)	CI (↑)	r_m^2 (↑)
Morgan (ECFP4)	0.181	0.897	0.765
MACCS (MolRes-DTA)	**0.168**	**0.906**	**0.778**

D. Effect of diverse fusion strategies with MolRes-DTA

We tested several fusion mechanisms under identical settings, including gating, cross-attention, and bilinear pooling, using the Davis dataset. As shown in Table S5, the concatenation-based fusion used in MolRes-DTA achieves the best overall performance, indicating its effectiveness in preserving complementary information from both modalities while maintaining model stability.

Table S5 Comparison of alternative fusion strategies

Dataset	Fusion Method	MSE (↓)	CI (↑)	r_m^2 (↑)
Davis	Gating	0.1771	0.8903	0.7904
	Cross-Attention	0.1813	0.8981	0.7658
	Bilinear	0.1827	0.9015	0.778
	MolRes-DTA	**0.1685**	**0.9065**	**0.7768**

E. Information complementarity analysis

(1) Visualization of feature complementarity.

We examined the mutual information between MACCS fingerprints and molecular graph embeddings to quantify feature independence and redundancy in Figure S1 (with mean MI = 0.12). The MI heatmap shows that most pairwise correlations fall within 20% (light yellow). The overall low MI indicates that the two representations encode largely independent and complementary molecular information.

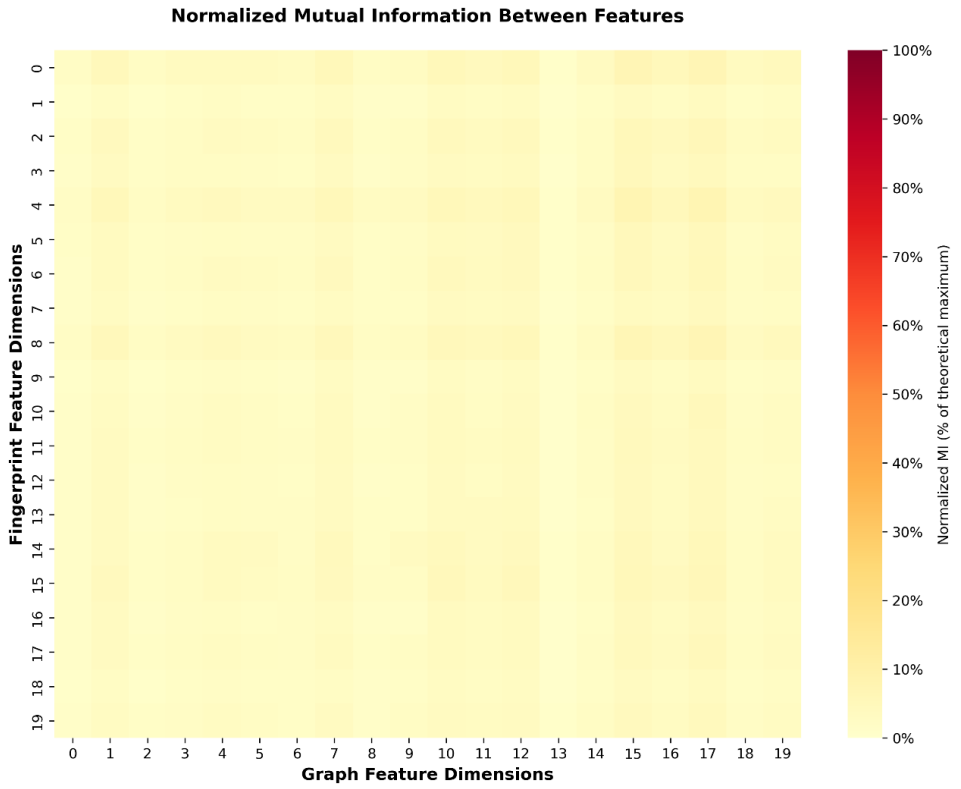

Fig. S1 Mutual information Heatmap of molecular fingerprint and molecular graph features.

(2) Per-group performance across molecular size strata.

To assess whether fusion contributes to robustness on larger molecules, we compared the prediction performance of fingerprint-only, graph-only, and fused models under different molecular size groups.

Table S6 Performance of fingerprint-only and fusion methods in DTA prediction by molecular size group

Molecular size	Model	MSE (↓)	CI (↑)
Small	Fingerprint Only	0.122	0.908
	Full Model (Fusion)	**0.082**	**0.912**
Medium	Fingerprint Only	0.182	0.893
	Full Model (Fusion)	**0.158**	**0.903**
Large	Fingerprint Only	0.220	0.880
	Full Model (Fusion)	**0.189**	**0.900**

F. Effect of drug size on affinity prediction

(1) Expanded dataset: KIBA SMILES-length analysis

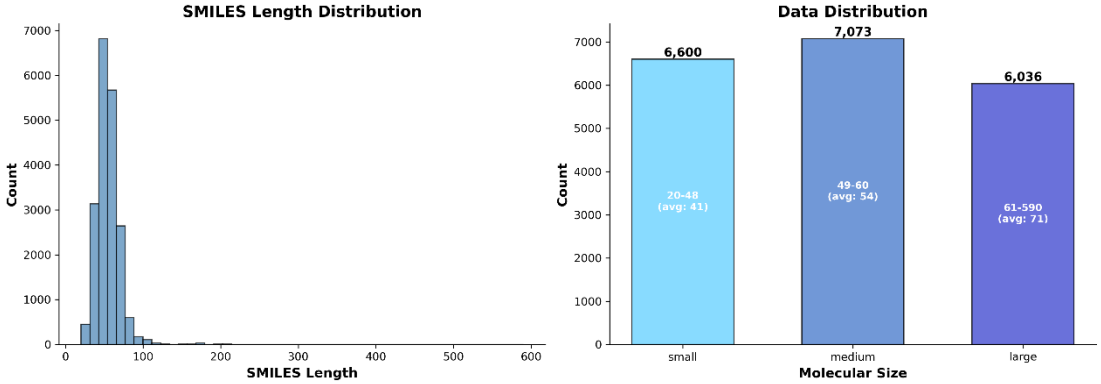

Fig. S2 The analysis of the KIBA dataset.

Table S7 KIBA test set (re-split by SMILES-length tertiles)

Group	MSE (↓)	CI (↑)	r_m^2 (↑)
small	0.118	0.894	0.758
medium	0.144	0.887	0.739
large	0.183	0.896	0.784

The KIBA results mirror the trend observed on the Davis dataset, where prediction error increases with increasing SMILES length. Because KIBA contains more diverse protein types, this concordance indicates that the effect of larger molecular size degrading model accuracy is not unique to kinase-only chemistry and has broader applicability.

(2) Extended Molecular-Size Evaluation.

To further examine the influence of molecular size on model performance, we computed three physicochemical descriptors for all compounds in the Davis test set: Molecular Surface Area (MSA), Van der Waals (VdW) volume, and Molecular Weight (MW). For each descriptor, the test set was partitioned into three equal-sized bins (small, medium, large) to ensure comparable sample sizes across bins. We then evaluated GraphDTA, TEFDTA, and MolRes-DTA on each bin and reported MSE, CI, and r_m^2 to

characterize performance variation across molecular-size classes. The detailed results are summarized in Table S8-S10.

Table S8 Davis results by molecular surface area

Model	Size	MSE (↓)	CI (↑)	r_m^2 (↑)
GraphDTA	small	0.151	0.873	0.155
	medium	0.212	0.899	0.344
	large	0.237	0.876	0.876
TEFDTA	small	0.141	0.889	0.659
	medium	0.265	0.897	0.782
	large	0.245	0.883	0.701
MolRes-DTA	small	**0.104**	**0.911**	**0.765**
	medium	**0.186**	**0.906**	**0.802**
	large	**0.171**	**0.903**	**0.752**

Table S9 Davis results by van der Waals volume

Model	Size	MSE (↓)	CI (↑)	r_m^2 (↑)
GraphDTA	small	0.179	0.882	0.248
	medium	0.219	0.88	0.273
	large	0.205	0.871	0.255
TEFDTA	small	0.187	0.895	0.728
	medium	0.230	0.887	0.725
	large	0.220	0.898	0.741
MolRes-DTA	small	**0.127**	**0.916**	**0.814**
	medium	**0.167**	**0.911**	**0.777**
	large	**0.163**	**0.916**	**0.779**

Table S10 Davis results by molecular weight

Model	Size	Count	MSE (↓)	CI (↑)	r_m^2 (↑)
GraphDTA	small	1926	0.163	0.877	0.196
	medium	1920	0.212	0.899	0.306
	large	1832	0.233	0.878	0.244
TEFDTA	small	1926	0.16	0.889	0.689
	medium	1920	0.226	**0.905**	0.757
	large	1832	0.248	0.876	0.717
MolRes-DTA	small	1926	**0.111**	**0.907**	**0.772**
	medium	1920	**0.168**	0.902	**0.781**
	large	1832	**0.181**	**0.907**	**0.775**

G. Molecular docking

Table S11 Comparison of predicted and actual affinities of drug-target pairs with different affinities and Docking results

Target	Ligand	Predict affinity	True affinity	Autodock vina
P12931 SRC	Dasatinib	9.655	9.677	-8.776
P36888 FLT3	Lestaurtinib	8.1875	8.267	-9.031
P04629 TRKA	Sorafenib	5.5459	5.201	-7.702

H. Molecular dynamics

All molecular dynamics simulations were performed with Amber24 and AmberTools25. Protein atoms were described by the ff19SB force field, water was modeled with the OPC explicit solvent, and the small-molecule ligand was parameterized with GAFF2 with AM1-BCC partial charges. Each protein–ligand complex was solvated in a cubic periodic box of OPC water, and Na^+ or Cl^- ions were added to neutralize the system. System setup, energy minimization and equilibration followed a standard multi-stage protocol implemented in AmberMdPrep.sh, involving restrained minimizations and short NVT/NPT runs with gradually released restraints on solute heavy atoms and backbone until the system density and thermodynamic properties were equilibrated. The equilibrated complexes were then subjected to 100 ns NPT production simulations at 300 K and 1 bar using a 2 fs time step, SHAKE constraints on bonds involving hydrogen, particle mesh Ewald treatment of long-range electrostatics with a conventional nonbonded cutoff, a Langevin thermostat for temperature control and isotropic pressure coupling for barostatting; trajectories were saved at regular intervals for subsequent analysis.

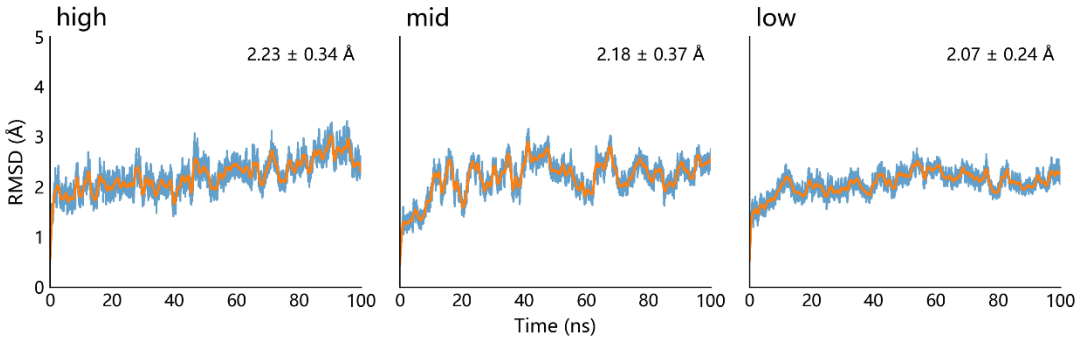

Fig. S3 Molecular dynamics simulations

During the 100 ns NPT production simulations all three protein–ligand complexes (**labelled high, mid and low according to their docking scores**) remained structurally stable. The complex RMSD increased from 0 to about 1.5–2.0 Å within the first few nanoseconds and then fluctuated around a nearly constant level for the remainder of the trajectories, indicating rapid relaxation from the docking pose followed by stable binding. The time-averaged RMSD values over the production runs were 2.23 ± 0.34 Å for **high**, 2.18 ± 0.37 Å for **mid** and 2.07 ± 0.24 Å for **low**, all around ~2 Å with moderate fluctuations. Thus, the overall protein fold is preserved and none of the ligands shows signs of dissociation or large-scale rearrangement on the 100 ns time scale.