Supplementary Materials

S1 Details of Protein Modifications and Domain Annotations

Table S1: Summary of protein modifications and domain annotations in the DAVIS dataset

Protein	Modification / Domain	Note
ABL1	E255K-phosphorylated	The phosphorylation site is Tyr393 [1].
	F317I	–
	F317I-phosphorylated	The phosphorylation site is Tyr393 [1].
	F317L	–
	F317L-phosphorylated	The phosphorylation site is Tyr393 [1].
	H396P	–
	H396P-phosphorylated	The phosphorylation site is Tyr393 [1].
	M351T-phosphorylated	The phosphorylation site is Tyr393 [1].
	Q252H	–
	Q252H-phosphorylated	The phosphorylation site is Tyr393 [1].
	T315I	–
	T315I-phosphorylated	The phosphorylation site is Tyr393 [1].
	Y253F-phosphorylated	The phosphorylation site is Tyr393 [1].
	Wild-type-phosphorylated	The phosphorylation site is Tyr393 [1].
BRAF	V600E	–
CDK4	CDK4-cyclinD1	CDK4-cyclinD1 complex
	CDK4-cyclinD3	CDK4-cyclinD3 complex
EGFR	E746A750del	–
	G719C	–
	G719S	–
	L747E749del, A750P	–
	L747E752del, P753S	–
	L747E751del, Sins	–
	L858R	–
	L858R, T790M	–
	L861Q	–
	S752I759del	–
	T790M	–
FGFR3	G697C	–
FLT3	D835H	–
	D835Y	–
	ITD	VDFREYEYDH insertion between Y591 and V592 [1]
	K663Q	–
	N841I	–
	R834Q	–
GCN2	Kinase domain 2, S808G	residues 590–1001 in UniProt Q9P2K8
JAK1	JH1 domain catalytic	residues 875-1153 in UniProt P23458 [2]
	JH2 domain pseudokinase	residues 583-855 in UniProt P23458 [2]
JAK2	JH1 domain catalytic	residues 849-1124 in UniProt O60674 [2]
JAK3	JH1 domain catalytic	residues 822-1111 in UniProt P52333 [2]
KIT	A829P	–
	D816H	–

Protein	Modification	Note
	D816V	–
	L576P	–
	V559D	–
	V559D, T670I	–
	V559D, V654A	–
LRRK2	G2019S	–
MET	M1250T	–
	Y1235D	–
PIK3CA	C420R	–
	E542K	–
	E545A	–
	E545K	–
	H1047L	–
	H1047Y	–
	I800L	–
	M1043I	–
	Q546K	–
RET	M918T	-
	V804L	–
	V804M	–
RPS6KA4	Kinase domain 1 N-terminal	residues 33-301 in UniProt O75676
	Kinase domain 2 C-terminal	residues 411-674 in UniProt O75676
RPS6KA5	Kinase domain 1 N-terminal	residues 49-318 in UniProt O75582
	Kinase domain 2 C-terminal	residues 426-687 in UniProt O75582
RSK1	Kinase domain 1 N-terminal	residues 62-321 in UniProt Q15418
	Kinase domain 2 C-terminal	residues 418-675 in UniProt Q15418
RSK2	Kinase domain 1 N-terminal	residues 68-327 in UniProt P51812
RSK3	Kinase domain 1 N-terminal	residues 59-318 in UniProt Q15349
	Kinase domain 2 C-terminal	residues 415-672 in UniProt Q15349
RSK4	Kinase domain 1 N-terminal	residues 73-330 in UniProt Q9UK32
	Kinase domain 2 C-terminal	residues 426-683 in UniProt Q9UK32
TYK2	JH1 domain catalytic	residues 897-1176 in UniProt P29597 [2]
	JH2 domain pseudokinase	residues 589-875 in UniProt P29597 [2]

S2 Binding Affinity Distribution

The DAVIS dataset contains 31,824 protein–ligand binding affinity measurements (442 proteins × 72 ligands). It is well known for its pronounced imbalance in binding affinity distribution: approximately 70% of protein–ligand pairs have K_d values capped at 10 μM (pK_d set to 5), leading to an overrepresentation of lower-affinity interactions [3]. The uncapped binding affinity distribution for all pairs is shown in Fig. S1(a). Among these, wild-type protein–ligand pairs include 20,126 capped measurements, with the corresponding uncapped distribution shown in Fig. S1(b). Modified protein–ligand pairs include 2,274 capped measurements, with the uncapped distribution shown in Fig. S1(c).

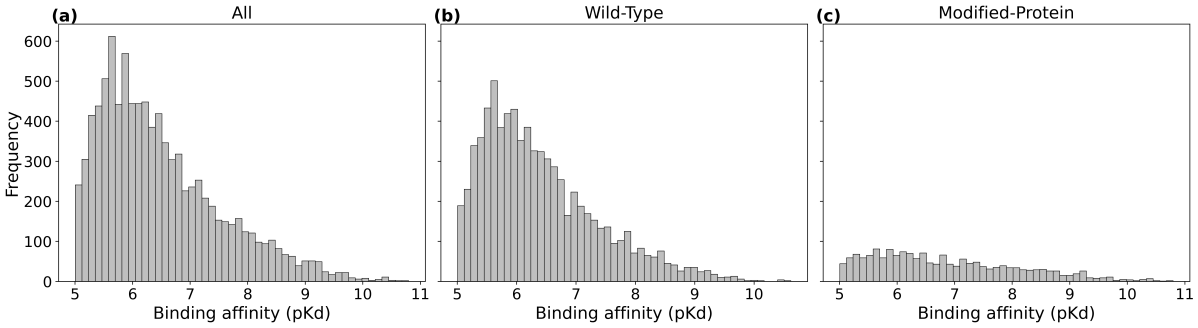

Figure S1: Distribution of uncapped binding affinity (pK_d) for (a) all protein–ligand pairs, (b) wild-type proteins, and (c) modified proteins

S3 Binding Affinity Alternation

To systematically evaluate how different protein modifications affect binding affinity, we performed a quantitative analysis and visualized the results as a heatmap (Fig. S2). The heatmap shows the magnitude of binding affinity changes induced by modifications across ABL1, BRAF, EGFR, FGFR3, FLT3, KIT, LRRK2, MET, PIK3CA, and RET. GCN2 has modified variants in the dataset, but its wild-type form is missing; therefore, affinity changes for GCN2 are not calculable and are excluded. The affinity change is defined as

$$\Delta pK_d = A(p_j^{m_i}, l_k) - A(p^{w_i}, l_k)$$

A key limitation of the DAVIS dataset is that binding affinity measurements (K_d) are capped at values above 10μM. As a result, we categorized modification-induced changes in binding affinity into four groups, as summarized in Table. S2 (1) WT-uncapped & modification-uncapped: both wild-type and modified proteins have K_d values below 10μM. The affinity changes are precisely trackable, and ΔpK_d reflects the exact magnitude of change. The distribution of these changes is shown in Fig. S3(a) (2) WT-capped & modification-uncapped: wild-type is capped ($K_d > 10$ μM), while the modified protein is not. This indicates an increase in affinity due to the modification. However, since the wild-type value is unknown beyond the threshold, ΔpK_d only represents a lower bound of the actual change.The distribution is shown in Fig. S3(b) (3) WT-uncapped & modification-capped: modified protein is capped, while the wild-type is not. This suggests a decrease in affinity, but again, ΔpK_d captures only the minimum possible magnitude, making the true change untrackable. The distribution is shown in Fig. S3(c) (4) WT-capped & modification-capped: both wild-type and modified proteins have capped affinities. In this case, the exact change in binding affinity is completely untrackable, and $\Delta pK_d = 0$

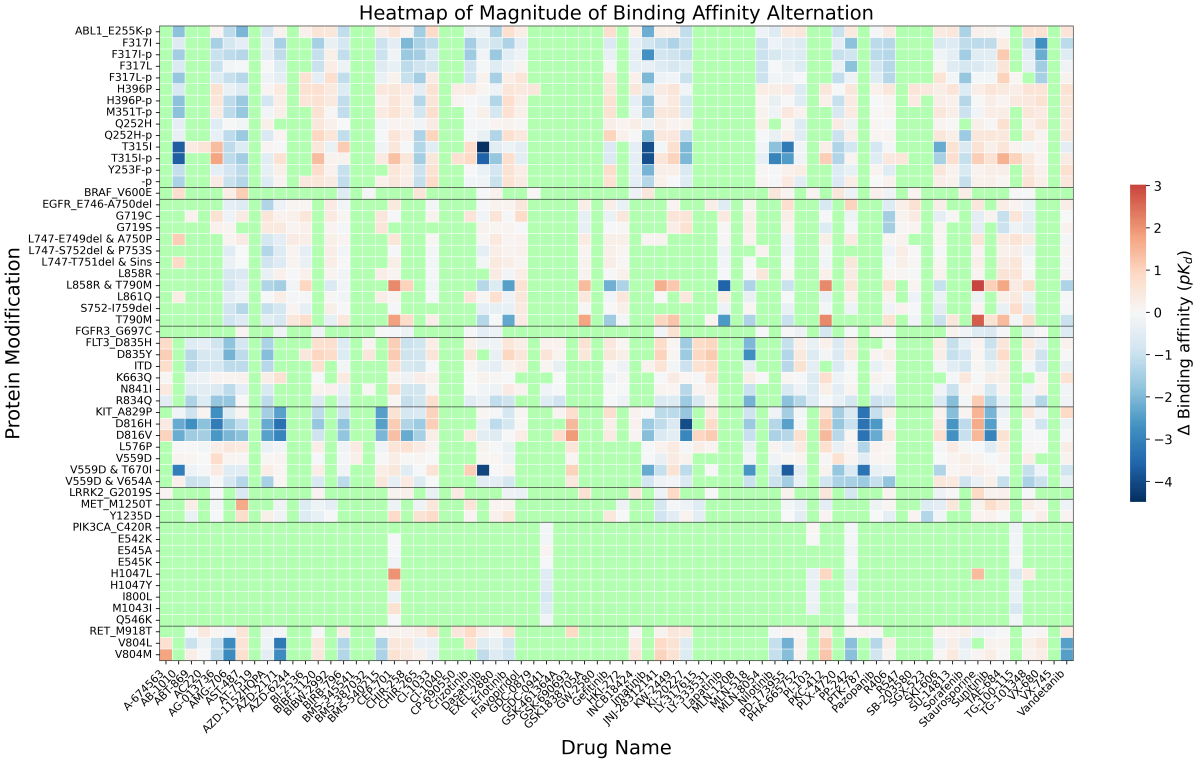

Figure S2: Heatmap of magnitude of binding affinity change. Colors from blue to red represent either the exact magnitude or the lower bound of the change, while light green indicates untrackable changes.

Table S2: Binding affinity summary between wild-type and modified proteins. ✓ indicates a protein-ligand pair with $K_d < 10$ μM, while ✗ indicates $K_d > 10$ μM.

WT	Modification	#
✓	✓	1601
✗	✓	134
✓	✗	157
✗	✗	2068

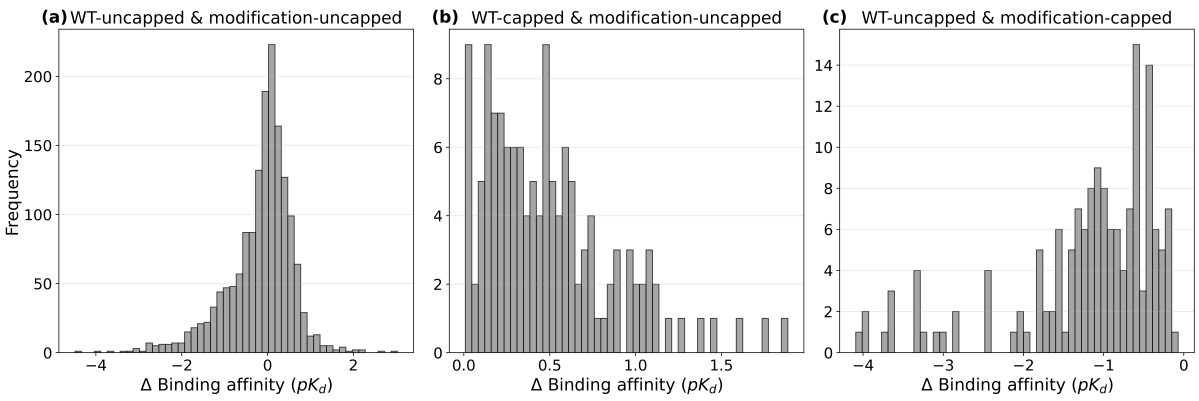

Figure S3: Distribution of magnitude of trackable binding affinity alternation. (a) Both wild-type and modified proteins have K_d values below 10μM. The affinity changes are precisely trackable. (b) Wild-type is capped ($K_d > 10$ μM), while the modified protein is not. (c) Modified protein is capped, while the wild-type is not.

S4 Input preprocessing

For the all benchmarks, the protein kinase domain is exclusively selected for each kinase protein. If domain annotations are absent in the DAVIS dataset, we utilize domain information from the UniProt database [4]. Phosphorylation events are not accounted for in docking-free methods due to their intrinsic input constraints; however, phosphorylated protein structures used as inputs for the FDA model are predicted using AlphaFold3 [5]. Similarly, AlphaFold3 is employed to predict structures of other modified proteins, whereas wild-type protein structures are directly sourced from the AlphaFold Protein Structure Database [6]. For the CDK4-cyclinD1 and CDK4-cyclinD3 complexes, amino acid sequences of both components are concatenated for docking-free model inputs, whereas their 3D structures are predicted using AlphaFold3 for the FDA model.

S5 Benchmark Models

We include 7 models (5 docking free-based models and 2 docking-based model) in the benchmarks.

DeepDTA DeepDTA [7] takes drug SMILES strings and protein amino acid sequences as input. It uses two parallel 1D CNNs to extract features from the drug and protein sequences, respectively. These features are then concatenated and passed through fully connected layers to predict the binding affinity. Our implementation is based on the code available at `https://github.com/KSUN63/DeepDTA-Pytorch`.

AttentionDTA AttentionDTA [8] takes SMILES strings and protein sequences as input but enhances DeepDTA by adding attention mechanisms. After initial feature extraction using 1D CNNs for both inputs, multi-head attention layers are applied to focus on important regions in the sequences. Our implementation is based on the code available at `https://github.com/zhaoqichang/AttentionDTA_TCBB`.

GraphDTA GraphDTA [9] represents the drug as a molecular graph (with atoms as nodes and bonds as edges, derived from the SMILES) and the protein as a sequence. A graph neural network (GCN and GAT) is used to process the drug graph, while a CNN processes the protein sequence. The learned representations are concatenated and passed to fully connected layers for affinity prediction. Our implementation is based on the code available at `https://github.com/thinng/GraphDTA`.

DGraphDTA DGraphDTA [10] encodes both the drug and the protein as graphs: drugs from molecular structures (via SMILES) and proteins from predicted contact maps. It applies GCNs to each graph separately, then combines the learned representations to predict the binding affinity. Our implementation is based on the code available at `https://github.com/595693085/DGraphDTA`.

MGraphDTA MGraphDTA [11] processes drug molecules as graphs and proteins as amino acid sequences. It employs a deep, multiscale graph neural network architecture with stacked GNN layers and dense skip connections to capture hierarchical structural features of the drug. Protein sequences are processed via 1D CNNs. The fused representations are used to predict binding affinity. Our implementation is based on the code available at `https://github.com/guaguabujianle/MGraphDTA`.

Folding-Docking-Affinity Folding-Docking-Affinity [12] uses protein sequences and drug molecular SMILES as input. The model operates in three stages: first, it predicts the 3D structure of the protein (e.g., via AlphaFold [5, 13]); second, it docks the drug molecule onto the predicted protein structure to generate a 3D complex through DiffDock [14]; third, it uses the 3D conformation of the protein-ligand complex as input to predict binding affinity through GIGN [15]. Notably, in our implementation, we ensemble five affinity predictors, and the final output is the mean of their predictions. Our implementation is based on the code available at `https://github.com/ZhiGroup/FDA`.

Boltz-2 Boltz-2 [16] is an open-source biomolecular foundation model that jointly predicts protein–ligand complex structures and binding affinities. In our runs, we follow the default affinity-inference settings: we generate 5 affinity diffusion samples, each with 200 reverse-diffusion steps; we then rank the resulting complexes by the protein–ligand pair ipTM score and feed the top-ranked pose to the affinity module. Our implementation is based on the code available at `https://github.com/jwohlwend/boltz`.

S6 Model Training Details

Table S3: Training hyperparameter settings for docking free-based models and Folding-Docking-Affinity (docking-based) used in Augmented Dataset Prediction and Wild-Type to Modification Generalization benchmarks.

Hyperparameter	**Docking free-based models**	**Folding-Docking-Affinity**
Optimizer	Adam	Adam
Learning rate	5e-4	5e-4
Weight decay	–	1e-6
Batch size	64	128
Max epochs	1,000	1,000
Early stopping patience	100	100

Table S4: Fine-tuning hyperparameter settings for docking free-based models and Folding-Docking-Affinity (docking-based) used in Few-Shot Modification Generalization.

Same-ligand, different-modifications		
Hyperparameter	**Docking free-based models**	**Folding-Docking-Affinity**
Optimizer	Adam	Adam
Learning rate	5e-4	5e-3
Weight decay	–	1e-6
Batch size	len(training set)	len(training set)
Number of epochs	30	30
Same-modification, different-ligands		
Optimizer	Adam	Adam
Learning rate	1e-4	1e-4
Weight decay	–	1e-6
Batch size	len(training set)	len(training set)
Number of epochs	10	10

S7 Metrics for Model Evaluation

We introduce two baseline methods for the same-ligand, different-modifications and same-modification, different-ligands settings: Wild-type ground truth (y_{WT}), which uses the ground truth binding affinity of the wild-type protein–ligand pair to predict that of the modified pairs, and wild-type prediction ($\hat{y}_{\mathrm{WT}}$), which uses the model-predicted affinity of the wild-type pair instead. We use both baselines to compute evaluation metrics—MSE, R_p, and C-index—for comparison with the predictions on the modified pairs. The following is the calculation details:

S7.1 Same-ligand, different-modifications

$$\mathrm{MSE}(y,\, y_{\mathrm{WT}}) = \frac{1}{n}\sum_{j=1}^{n}(A(p^{w_i}, l_k) - A(p_j^{m_i}, l_k))^2$$

where the ground-truth binding affinity of the wild-type kinase–ligand pair, $A(p^{w_i}, l_k)$, is used to compute Mean Squared Error, denoted as $\mathrm{MSE}(y,\, y_{\mathrm{WT}})$.

$$\mathrm{MSE}(y,\, \hat{y}_{\mathrm{WT}}) = \frac{1}{n}\sum_{j=1}^{n}(f(p^{w_i}, l_k) - A(p_j^{m_i}, l_k))^2$$

where the model-predicted binding affinity for the wild-type kinase–ligand pair, $f(p^{w_i}, l_k)$, is used to compute Mean Squared Error, denoted as $\mathrm{MSE}(y,\, \hat{y}_{\mathrm{WT}})$.

$$\mathrm{MSE} = \frac{1}{n}\sum_{j=1}^{n}(f(p_j^{m_i}, l_k) - A(p_j^{m_i}, l_k))^2$$

where the model-predicted binding affinity for the modified kinase–ligand pairs, $f(p_j^{m_i}, l_k)$, is used to compute Mean Squared Error, denoted as MSE.

$$R_p = \frac{\sum_{j=1}^{n}(f(p_j^{m_i}, l_k) - \overline{f(p_j^{m_i}, l_k)})(A(p_j^{m_i}, l_k) - \overline{A(p_j^{m_i}, l_k)})}{\sqrt{\sum_{j=1}^{n}(f(p_j^{m_i}, l_k) - \overline{f(p_j^{m_i}, l_k)})^2}\sqrt{\sum_{j=1}^{n}(A(p_j^{m_i}, l_k) - \overline{A(p_j^{m_i}, l_k)})^2}}$$

where the model-predicted binding affinity, $f(p_j^{m_i}, l_k)$, and ground-truth binding affinity, $A(p_j^{m_i}, l_k)$, for the modified kinase–ligand pairs are used to compute Pearson correlation coefficient, R_p.

$$\text{C-index} = \frac{\sum_{r \neq s} I(A(p_r^{m_i}, l_k) > A(p_s^{m_i}, l_k)) \cdot I(f(p_r^{m_i}, l_k) > f(p_s^{m_i}, l_k)) + 0.5 \cdot I(f(p_r^{m_i}, l_k) = f(p_s^{m_i}, l_k))}{\sum_{r \neq s} I(A(p_r^{m_i}, l_k) > A(p_s^{m_i}, l_k))}$$

where the model-predicted binding affinity, $f(p_{r/s}^{m_i}, l_k)$, and ground-truth binding affinity, $A(p_{r/s}^{m_i}, l_k)$, for the modified kinase–ligand pairs are used to compute C-index. $I(\cdot)$ is denoted as an indicator function that returns 1 if the condition is true and 0 otherwise

S7.2 Same-modification, different-ligands

$$\text{MSE}(y,\, y_{\text{WT}}) = \frac{1}{n}\sum_{k=1}^{n}(A(p^{w_i}, l_k) - A(p_j^{m_i}, l_k))^2$$

where the ground-truth binding affinity of the wild-type kinase–ligand pairs, $A(p^{w_i}, l_k)$, is used to compute Mean Squared Error.

$$\text{MSE}(y,\, \hat{y}_{\text{WT}}) = \frac{1}{n}\sum_{k=1}^{n}(f(p^{w_i}, l_k) - A(p_j^{m_i}, l_k))^2$$

where the model-predicted binding affinity for the wild-type kinase–ligand pairs, $f(p^{w_i}, l_k)$, is used to compute Mean Squared Error.

$$\text{MSE} = \frac{1}{n}\sum_{k=1}^{n}(f(p_j^{m_i}, l_k) - A(p_j^{m_i}, l_k))^2$$

where the model-predicted binding affinity for the modified kinase–ligand pairs, $f(p_j^{m_i}, l_k)$, is used to compute Mean Squared Error.

$$R_p(y,\, y_{\text{WT}}) = \frac{\sum_{k=1}^{n}(A(p^{w_i}, l_k) - \overline{A(p^{w_i}, l_k)})(A(p_j^{m_i}, l_k) - \overline{A(p_j^{m_i}, l_k)})}{\sqrt{\sum_{k=1}^{n}(A(p^{w_i}, l_k) - \overline{A(p^{w_i}, l_k)})^2}\sqrt{\sum_{k=1}^{n}(A(p_j^{m_i}, l_k) - \overline{A(p_j^{m_i}, l_k)})^2}}$$

where the ground-truth binding affinity, $A(p^{w_i}, l_k)$, for the wild-type kinase–ligand pairs and the ground-truth binding affinity, $A(p_j^{m_i}, l_k)$, for the modified kinase–ligand pairs are used to compute Pearson correlation coefficient.

$$R_p(y,\, \hat{y}_{\text{WT}}) = \frac{\sum_{k=1}^{n}(f(p^{w_i}, l_k) - \overline{f(p^{w_i}, l_k)})(A(p_j^{m_i}, l_k) - \overline{A(p_j^{m_i}, l_k)})}{\sqrt{\sum_{k=1}^{n}(f(p^{w_i}, l_k) - \overline{f(p^{w_i}, l_k)})^2}\sqrt{\sum_{k=1}^{n}(A(p_j^{m_i}, l_k) - \overline{A(p_j^{m_i}, l_k)})^2}}$$

where the model-predicted binding affinity, $f(p^{w_i}, l_k)$, for the wild-type kinase–ligand pairs and the ground-truth binding affinity, $A(p_j^{m_i}, l_k)$, for the modified kinase–ligand pairs are used to compute Pearson correlation coefficient.

$$R_p = \frac{\sum_{k=1}^{n}(f(p_j^{m_i}, l_k) - \overline{f(p_j^{m_i}, l_k)})(A(p_j^{m_i}, l_k) - \overline{A(p_j^{m_i}, l_k)})}{\sqrt{\sum_{k=1}^{n}(f(p_j^{m_i}, l_k) - \overline{f(p_j^{m_i}, l_k)})^2}\sqrt{\sum_{k=1}^{n}(A(p_j^{m_i}, l_k) - \overline{A(p_j^{m_i}, l_k)})^2}}$$

where the model-predicted binding affinity, $f(p_j^{m_i}, l_k)$, and ground-truth binding affinity, $A(p_j^{m_i}, l_k)$, for the modified kinase–ligand pairs are used to compute Pearson correlation coefficient, R_p.

$$\text{C-index}(y,\, y_{\text{WT}}) = \frac{\sum_{r \neq s} I(A(p_j^{m_i}, l_r) > A(p_j^{m_i}, l_s)) \cdot I(A(p^{w_i}, l_r) > A(p^{w_i}, l_s)) + 0.5 \cdot I(A(p^{w_i}, l_r) = A(p^{w_i}, l_s))}{\sum_{r \neq s} I(A(p_j^{m_i}, l_r) > A(p_j^{m_i}, l_s))}$$

where the ground-truth binding affinity, $A(p^{w_i}, l_{r/s})$, for the wild-type kinase–ligand pairs and the ground-truth binding affinity, $A(p_j^{m_i}, l_{r/s})$, for the modified kinase–ligand pairs are used to compute concordance index.

$$\text{C-index}(y,\, \hat{y}_{\text{WT}}) = \frac{\sum_{r \neq s} I(A(p_j^{m_i}, l_r) > A(p_j^{m_i}, l_s)) \cdot I(f(p^{w_i}, l_r) > f(p^{w_i}, l_s)) + 0.5 \cdot I(f(p^{w_i}, l_r) = f(p^{w_i}, l_s))}{\sum_{r \neq s} I(A(p_j^{m_i}, l_r) > A(p_j^{m_i}, l_s))}$$

where the model-predicted binding affinity, $f(p^{w_i}, l_{r/s})$, for the wild-type kinase–ligand pairs and the ground-truth binding affinity, $A(p_j^{m_i}, l_{r/s})$, for the modified kinase–ligand pairs are used to compute concordance index.

$$\text{C-index} = \frac{\sum_{r \neq s} I(A(p_j^{m_i}, l_r) > A(p_j^{m_i}, l_s)) \cdot I(f(p_j^{m_i}, l_r) > f(p_j^{m_i}, l_s)) + 0.5 \cdot I(f(p_j^{m_i}, l_r) = f(p_j^{m_i}, l_s))}{\sum_{r \neq s} I(A(p_j^{m_i}, l_r) > A(p_j^{m_i}, l_s))}$$

where the model-predicted binding affinity, $f(p_j^{m_i}, l_{r/s})$, and ground-truth binding affinity, $A(p_j^{m_i}, l_{r/s})$, for the modified kinase–ligand pairs are used to compute concordance index.

S8 Supplementary figures

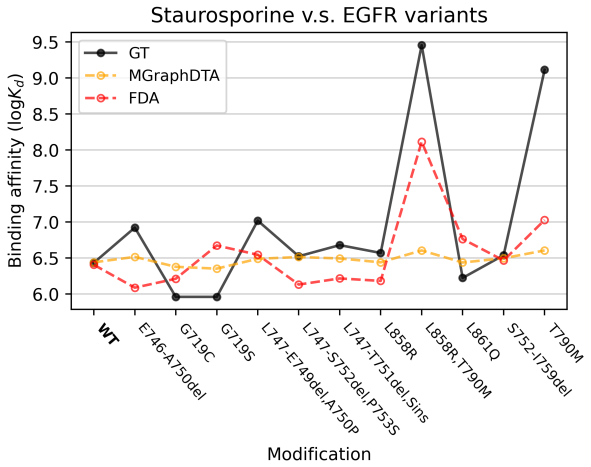

Figure S4: A case of the Wild-Type to Modification Generalization benchmark: Same-ligand, different-modifications — Staurosporine binding to various EGFR protein variants.

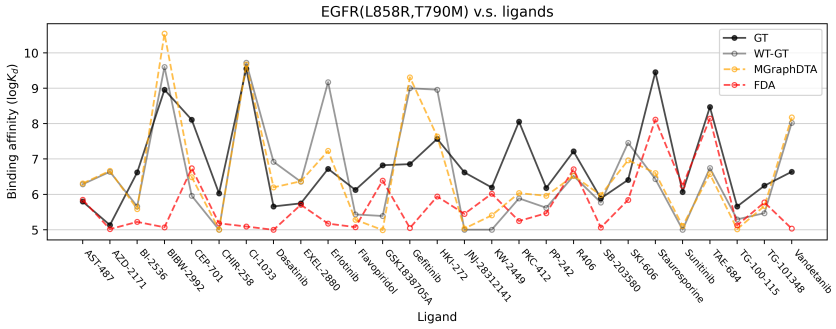

Figure S5: A case of the Wild-Type to Modification Generalization benchmark: Same-modification, different-ligands — the EGFR(L858R, T790M) variant interacting with multiple ligands.

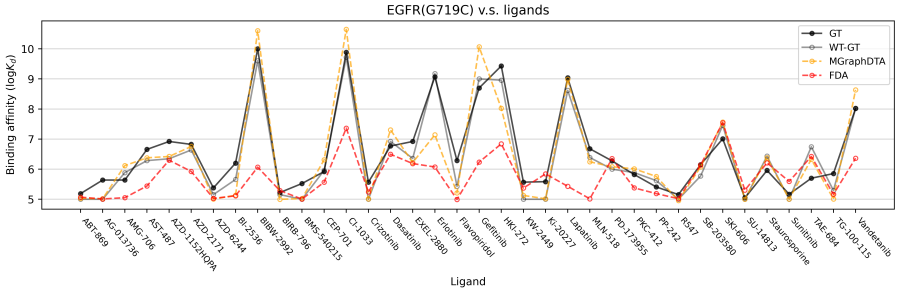

Figure S6: A case of the Wild-Type to Modification Generalization benchmark: Same-modification, different-ligands — the EGFR(G719C) variant interacting with multiple ligands.

References

[1] Lisa M Wodicka, Pietro Ciceri, Mindy I Davis, Jeremy P Hunt, Mark Floyd, Sara Salerno, Xuequn H Hua, Julia M Ford, Robert C Armstrong, Patrick P Zarrinkar, et al. Activation state-dependent binding of small molecule kinase inhibitors: structural insights from biochemistry. *Chemistry & biology*, 17(11):1241–1249, 2010.

[2] Xiaoyi Hu, Jing Li, Maorong Fu, Xia Zhao, and Wei Wang. The jak/stat signaling pathway: from bench to clinic. *Signal transduction and targeted therapy*, 6(1):402, 2021.

[3] 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*, 29:1046–1051, 11 2011.

[4] Uniprot: the universal protein knowledgebase in 2025. *Nucleic Acids Research*, 53(D1):D609–D617, 2025.

[5] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pages 1–3, 2024.

[6] Mihaly Varadi, Damian Bertoni, Paulyna Magana, Urmila Paramval, Ivanna Pidruchna, Malarvizhi Radhakrishnan, Maxim Tsenkov, Sreenath Nair, Milot Mirdita, Jingi Yeo, et al. Alphafold protein structure database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic acids research*, 52(D1):D368–D375, 2024.

[7] Hakime Öztürk, Arzucan Özgür, and Elif Ozkirimli. Deepdta: Deep drug-target binding affinity prediction. *Bioinformatics*, 34:i821–i829, 9 2018.

[8] Qichang Zhao, Guihua Duan, Mengyun Yang, Zhongjian Cheng, Yaohang Li, and Jianxin Wang. Attentiondta: Drug–target binding affinity prediction by sequence-based deep learning with attention mechanism. *IEEE/ACM transactions on computational biology and bioinformatics*, 20(2):852–863, 2022.

[9] 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.

[10] 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.

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

[12] Ming-Hsiu Wu, Ziqian Xie, and Degui Zhi. A folding-docking-affinity framework for protein-ligand binding affinity prediction. *Communications Chemistry*, 8(1):1–9, 2025.

[13] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, et al. Protein complex prediction with alphafold-multimer. *biorxiv*, pages 2021–10, 2021.

[14] Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

[15] Ziduo Yang, Weihe Zhong, Qiujie Lv, Tiejun Dong, and Calvin Yu-Chian Chen. Geometric interaction graph neural network for predicting protein-ligand binding affinities from 3d structures (gign). *The Journal of Physical Chemistry Letters*, 14(8):2020–2033, 2023.

[16] Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, et al. Boltz-2: Towards accurate and efficient binding affinity prediction. *BioRxiv*, pages 2025–06, 2025.