NO LIGAND BINDS IN ISOLATION: COMPETITION-AWARE BIOACTIVITY PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate prediction of drug–target bioactivity is central to early-stage drug discovery, yet existing deep learning models evaluate protein–ligand interactions in isolation, ignoring the competitive binding context that governs drug action in cellular environments. In radioligand displacement assays, observed bioactivity is modulated by the relative binding strength and spatial pose overlap between a test ligand and a competing radioligand within the same protein pocket—factors that conventional single-ligand models neglect. We present DrugComp, a competition-aware deep learning framework that explicitly encodes atom-level binding affinities and steric overlaps between competing ligands within shared protein binding pockets. DrugComp integrates seamlessly with diverse backbone architectures, including graph neural networks and transformers. To support competition-aware learning, we curate CompDB, a dataset of 10,675 radioligand displacement assays spanning 151 proteins, 181 radioligands, and 8,674 test ligands, enriched with docking-derived structural context. Evaluated in five-fold cross-validation, DrugComp consistently outperforms competition-blind baselines across all architectures, achieving up to 44.7% improvement in bioactivity prediction metrics. These results demonstrate that reframing bioactivity prediction as competitive-context modeling substantially improves predictive performance and better reflects the mechanistic basis of drug action.

1 INTRODUCTION

Drug–target bioactivity quantifies the extent to which a compound engages a biological target and modulates its functional outcome. Its accurate prediction is a central problem in early-stage drug discovery, where the goal is to identify potential drugs that can interact with and regulate target proteins before costly experimental validation. Accurate prediction of drug–target bioactivity is central to virtual screening, lead optimization, and the quantitative characterization of molecular efficacy Zhou et al. (2025); Sadybekov & Katritch (2023); Yang et al. (2026). By estimating molecular bioactivity at scale, computational bioactivity models can help prioritize promising compounds from vast chemical spaces and guide the discovery of candidates with desirable bioactivity and selectivity.

Recently, deep learning (DL)-based approaches have garnered considerable attention in bioactivity prediction and demonstrated impressive capabilities in modeling complex drug–target interactions Yang et al. (2023); Yin et al. (2024); Wang et al. (2024); He et al. (2026); Wang et al. (2025a); Kyro et al. (2025); Lim et al. (2025); Wang et al. (2025b); Li & Gong (2025); Wu et al. (2026). Despite these advances, existing models remain centered on isolated drug–target pairs. Yet, in cells, drug bioactivity is not governed solely by affinity for a purified target, but by competition with endogenous ligands that occupy the same binding sites. Ligands rarely act alone in biological systems, as they coexist within a shared molecular environment, where relative affinity, concentration, and spatial compatibility jointly determine target occupancy and apparent bioactivity Martinez-Hackert et al. (2021). This competitive context is fundamental to cellular pharmacology, but remains beyond the scope of existing DL-based models, which cannot explicitly capture how endogenous or other ligands reshape apparent bioactivity. Although high target affinity is often sought in drug discovery, affinity alone does not guarantee cellular efficacy: a ligand that binds strongly to a purified target may still be ineffective if it fails to outcompete endogenous ligands

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.

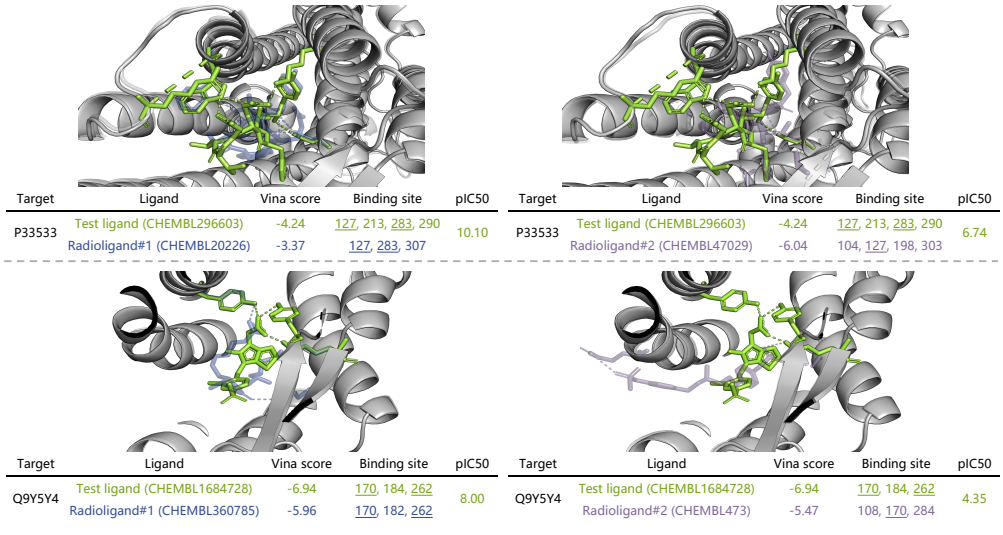


Figure 1: Bioactivity is shaped by competitive binding context. (Top) Relative binding strength modulates bioactivity. For target P33533, the same test ligand exhibits high bioactivity against a weaker radioligand but reduced bioactivity against a stronger radioligand. (Bottom) Pose overlap determines competitive displacement. For target Q9Y5Y4, extensive pose overlap leads to stronger displacement, whereas partial overlap limits inhibition. Docking poses are generated with AutoDock Vina Trott & Olson (2010).

in the same pocket. The relevant question is therefore not only whether a compound binds, but whether it binds strongly enough, and in a compatible pose, to win the competition.

Biochemical radioligand displacement assays make this limitation particularly apparent Maguire et al. (2012). In this assay, the target protein is first incubated with a radioligand, often chosen as a proxy for the endogenous ligand or its analog, allowing the radioligand to occupy the binding site. The test ligand is then titrated into the system, where it competes with the radioligand for the same binding site. The measured signal therefore reports a competitive displacement process, rather than the intrinsic bioactivity of an isolated protein–ligand pair. Accordingly, the observed bioactivity is intrinsically context-dependent: for a fixed protein target and test ligand, measured bioactivity can shift substantially depending on which radioligand is used. As illustrated in Fig. 1, this dependency arises from two coupled determinants: (*i*) the relative binding strength of the competing ligands—a stronger radioligand can markedly suppress the apparent bioactivity of a test ligand; and (*ii*) the spatial overlap of their binding poses within the shared pocket, where limited geometric overlap reduces direct competition and lowers observed inhibition even at comparable affinities.

As shown in Fig. 1 (Top), for the delta-type opioid receptor (UniProt ID: P33533), the test ligand CHEMBL296603 (Deltorphine II) exhibits high experimentally measured bioactivity when competing with the weaker-affinity radioligand#1 CHEMBL20226 (Deltorphin II), with a pIC_{50} of 10.10 Fichna et al. (2011). However, when the same test ligand is assayed against the stronger-affinity radioligand#2 CHEMBL47029 (DSLET), the experimentally measured bioactivity drops to a pIC_{50} of 6.74 Salvadori et al. (1991), indicating that apparent potency can be substantially modulated by the relative binding strength of the radioligand. In addition, geometric overlap provides a second determinant of competitive bioactivity, as shown in Fig. 1 (Bottom). For Prostaglandin D2 receptor 2 (UniProt ID: Q9Y5Y4), the test ligand CHEMBL1684728 substantially overlaps with radioligand#1 CHEMBL360785 ([^{3}H]PGD2) in the binding pocket, consistent with strong inhibition, with a pIC_{50} of 8.00 Bonafoux et al. (2011). In contrast, radioligand#2 CHEMBL473 (Dofetilide) occupies a partially distinct region near the pocket entrance, limiting direct displacement and leading to weaker inhibition, with a pIC_{50} of 4.35 Bonafoux et al. (2011), despite comparable predicted affinities. However, conventional models ignore the radioligand and collapse distinct competitive measurements into a single protein–ligand label, introducing ambiguity into bioactivity prediction.

To address this gap, we present DrugComp, a competition-aware DL framework for bioactivity prediction. Rather than treating each protein–ligand pair as an isolated observation, DrugComp learns directly from radioligand displacement assays, where measured bioactivity arises from

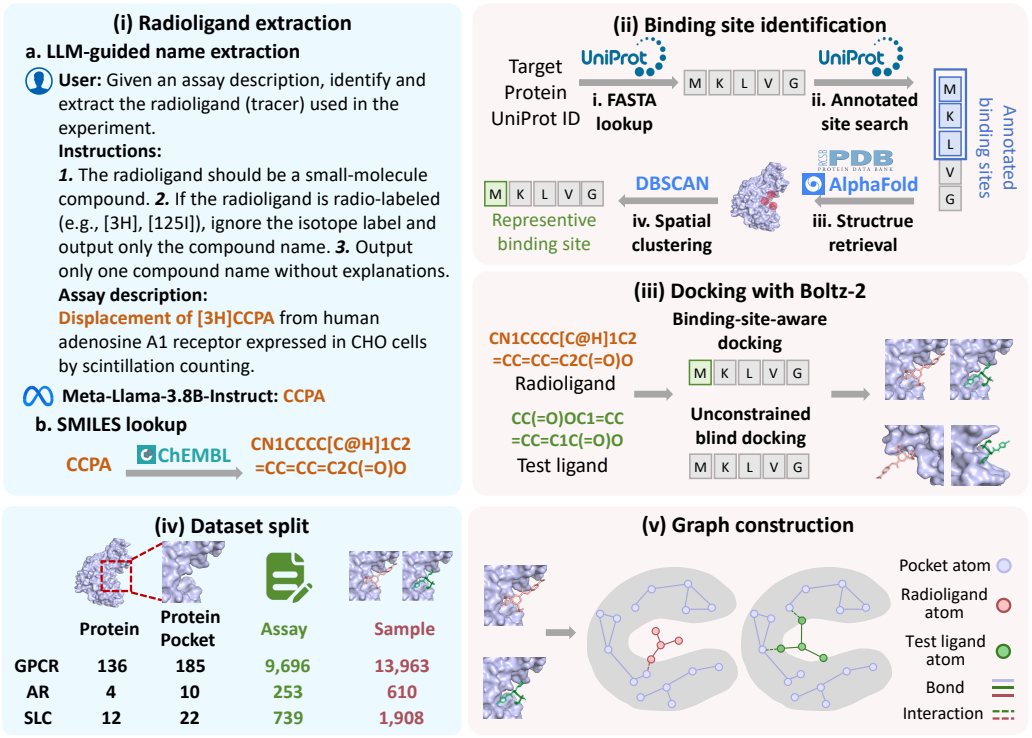


Figure 2: Construction workflow of the CompDB dataset. (i) We first use LLM to extract radioligand identities from assay descriptions and map them to canonical SMILES. (ii) Protein structures are retrieved from PDB or AlphaFold, and known binding sites are clustered to define representative docking targets. (iii) We then independently dock test ligands and radioligands with Boltz-2. (iv) Docked complexes are grouped by protein family, including GPCR, AR, and SLC, following AssayCTX. (v) Protein–ligand complexes are finally converted into graphs following DTIGN.

competition between a test ligand and a radioligand within the same binding pocket. The framework captures two complementary determinants of competitive bioactivity: the relative binding strength of the competing ligands and the spatial overlap of their preferred binding poses. By integrating these signals, DrugComp estimates how competitive ligands reshape the apparent bioactivity of candidate molecules across diverse protein targets. To support competition-aware learning, we curate CompDB, a dataset of 10,675 radioligand displacement measurements spanning 151 protein targets, 181 radioligands, and 8,674 test ligands, enriched with docking-derived structural context. In five-fold cross-validation on CompDB, DrugComp consistently improves seven competition-blind baselines, with gains of up to 44.7% in competition-adjusted bioactivity prediction. Replacing the assayed radioligand with a different radioligand of the same target costs 3.5–6.3% RMSE, and replacing it with an arbitrary ligand costs 7.1–8.1%. The advantage holds under ligand-, scaffold-, and target-disjoint evaluation. These results suggest that bioactivity prediction can be reframed from isolated scoring to competitive modeling, providing a path towards AI systems that better reflect how drugs act in cellular environments.

2 CompDB Dataset Construction

To enable competition-aware learning, we curate CompDB. To the best of our knowledge, CompDB is the first dataset that explicitly links test ligands to competing radioligands within shared protein binding environments. CompDB is built upon the publicly available AssayCTX resource Schoenmaker et al. (2025), which provides test ligand SMILES, target UniProt identifiers, harmonized assay metadata, and experimentally measured bioactivity. From AssayCTX, we select radioligand displacement assays Maguire et al. (2012), where a test ligand competes with a pre-bound radioligand for target binding. Since AssayCTX does not provide 3D protein–

ligand complexes, we supplement each assay with structural context by docking test ligands and radioligands separately to their target proteins using Boltz-2 Passaro et al. (2025). Docking is performed in blind mode by default using ligand SMILES and protein sequences. When binding-site annotations are available, we further perform binding-site-aware docking by restricting the search space to the specified residues. Therefore, we retrieve radioligand SMILES, target protein sequences, and binding-site residues for each assay. Together, these steps enrich assay-level competitive bioactivity measurements with structural information in the shared binding site. The detailed workflow is described below and illustrated in Fig. 2.

Radioligand Extraction. We first parse radioligand displacement assay descriptions from the original AssayCTX dataset using Meta-Llama-3.8B-Instruct Grattafiori et al. (2024) to extract radioligand identities. To validate the extraction, we manually review 100 randomly sampled assays and confirm that all extracted ligand names are correct. We then map the extracted names to canonical SMILES using the ChEMBL API Gaulton et al. (2012). After retaining radioligand displacement assays with successfully extracted radioligands and IC_{50}-type measurements, CompDB contains 181 unique radioligands, 151 unique proteins, and 8,674 unique test ligands.

Binding Site Identification. Given the UniProt accession number of each target protein Consortium (2015), we retrieve the corresponding amino acid sequence and known binding-site annotations from UniProt. For proteins with known binding sites, we obtain their spatial coordinates from experimentally resolved structures in the RCSB Protein Data Bank (PDB) Rose et al. (2016). When no experimental structure is available, we use AlphaFold3-predicted structures Abramson et al. (2024). We then cluster their known binding sites using DBSCAN Deng (2020) with a distance threshold of 12Å, yielding an average of 1.52 binding-site clusters per protein across the 42 proteins with known binding sites. For each cluster, we select the residue closest to the cluster centroid as the representative binding site for subsequent docking.

Docking. All docking operations are performed using Boltz-2 Passaro et al. (2025), a state-of-the-art deep learning-based protein–ligand co-folding model. For proteins with known binding sites, both test ligands and radioligands are independently docked at each representative binding site, with the search space constrained using the default distance threshold of 6 Å. Blind docking is also performed to explore alternative binding regions. For proteins without known binding sites, only blind docking is performed. This procedure yields 16,481 protein–test-ligand complexes and 505 protein–radioligand complexes across 151 proteins.

Dataset Split. Following the original AssayCTX split protocol Schoenmaker et al. (2025), we organize the datasets by protein family: G protein-coupled receptors (GPCRs), adenosine receptors (ARs; a GPCR subfamily used for ablation), and solute carriers (SLCs). The resulting CompDB-GPCR, CompDB-AR, and CompDB-SLC datasets comprise 13,963, 610, and 1,908 protein–ligand complexes, corresponding to 9,696, 253, and 739 radioligand displacement assays, respectively. Detailed statistics are provided in Appendix Table 7.

Graph Construction. Following DTIGN Yin et al. (2024), we construct drug–target interaction graphs with a slightly enlarged pocket cutoff. Instead of selecting pocket atoms within 5 Å of the ligand, we include atoms within 6 Å, consistent with Boltz-2's definition of atomic contacts. We also adopt the Union-Pocket representation from LigoSpace Zhou et al. (2025), which aggregates pocket atoms across multiple ligand poses to provide richer structural context. Specifically, we construct one Union-Pocket for each representative binding site and one additional Union-Pocket for blind docking regions, yielding 214 Union-Pockets in total. Since Boltz-2 may predict complex conformations in different coordinate frames, we align all coordinates before graph construction.

3 METHODS

3.1 COMPETITIVE BINDING GRAPH

The observed bioactivity of a ligand is governed not only by its intrinsic properties with respect to the protein, but also by its relative competitiveness against other coexisting ligands. To model this effect, we construct a competitive binding graph that jointly encodes the protein pocket, the test ligand, and the radioligand, as shown in Fig. 3. Specifically, during dataset construction, both the

test ligand and radioligand are docked into the same protein structure. We then standardize their spatial context by mapping both ligand poses onto a unified coordinate frame based on the aligned binding pocket. This yields a consistent graph representation in which different ligand poses are comparable within a shared structural context.

Formally, let L, R, and P denote the test ligand, radioligand, and protein, respectively. We define the competitive triplet graph as

$$\mathcal{G}(L, R, P) = (\mathcal{V}, \mathcal{E}, \mathcal{R}), \tag{1}$$

where the node set, edge set, and coordinate set are given by

$$\begin{aligned} \mathcal{V} &= \mathcal{V}_l \cup \mathcal{V}_r \cup \mathcal{V}_p, \\ \mathcal{E} &= \mathcal{E}_l \cup \mathcal{E}_r \cup \mathcal{E}_p \cup \mathcal{E}_{lp} \cup \mathcal{E}_{rp}, \\ \mathcal{R} &= \mathcal{R}_l \cup \mathcal{R}_r \cup \mathcal{R}_p. \end{aligned} \tag{2}$$

Here, $\mathcal{V}_l$, $\mathcal{V}_r$, and $\mathcal{V}_p$ denote the atoms of the test ligand, radioligand, and protein, respectively. The intra-molecular edge sets $\mathcal{E}_l$, $\mathcal{E}_r$, and $\mathcal{E}_p$ encode covalent bonds within each component. The inter-molecular edge sets $\mathcal{E}_{lp}$ and $\mathcal{E}_{rp}$ capture binding interactions between the protein and the test ligand or radioligand, respectively. Finally, $\mathcal{R}_l$, $\mathcal{R}_r$, and $\mathcal{R}_p$ represent the corresponding 3D atomic coordinates of the test ligand, radioligand, and protein.

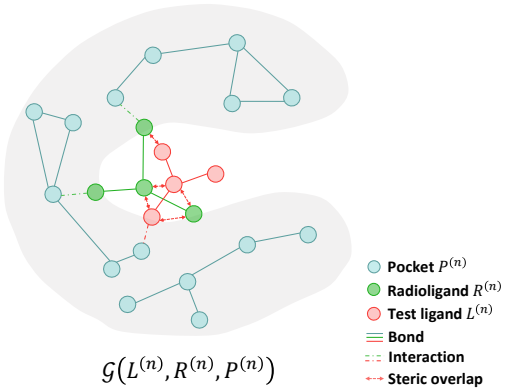


Figure 3: Competitive binding graph representation, with pocket, test-ligand, and radioligand atoms integrated into a unified graph.

Each node $v_i \in \mathcal{V}$ is associated with both chemical and geometric information, with the exact feature definitions following the original implementations of each baseline model, as detailed in Appendix Table 9. Common chemical features include node type, atom type, number of covalent bonds, implicit valence, hybridization state, aromaticity, and the number of attached hydrogen atoms. In addition, each node is assigned a 3D coordinate $\mathbf{r}_i \in \mathcal{R}$, which specifies its spatial position in the binding complex. For intra-molecular connectivity, an edge $e_{ij} \in \mathcal{E}_l \cup \mathcal{E}_r \cup \mathcal{E}_p$ is created between two atoms v_i and v_j when a covalent bond exists between them, whose features are also defined in a model-dependent manner, as summarized in Appendix Table 10. Common bond-level features include bond type, stereochemistry, conjugation, and ring membership. For inter-molecular interactions, an edge $e_{ij} \in \mathcal{E}_{lp} \cup \mathcal{E}_{rp}$ is added when the Euclidean distance $d_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|_2$ between atoms v_i and v_j is smaller than 6 Å. The detailed definitions of graph features are provided in Appendix F.2.

Once the competitive triplet graph $\mathcal{G}$ is constructed, the learning objective is to predict the bioactivity y from the graph representation. This can be formulated as $F : \mathcal{G} \to y$, where F denotes the graph-based prediction model. In practice, a protein may exhibit multiple conformations and contain multiple potential binding pockets. Therefore, for a given test ligand–radioligand–protein triplet, we may obtain a set of N competitive triplet graphs, each corresponding to a representative protein conformation or binding site. For the n-th graph, the predicted bioactivity is given by

$$\hat{y}^{(n)} = F\left(\mathcal{G}^{(n)}\right), \tag{3}$$

where $\mathcal{G}^{(n)} = \mathcal{G}(L^{(n)}, R^{(n)}, P^{(n)})$ denotes the competitive graph constructed from the n-th binding configuration. To improve prediction robustness, we adopt an ensemble strategy by averaging the predictions over all graphs:

$$\hat{y}_{\text{final}} = \frac{1}{N} \sum_{n=1}^{N} \hat{y}^{(n)}, \tag{4}$$

which allows the model to account for conformational and pocket-level variability, thereby improving predictive stability across diverse protein–ligand complexes.

3.2 Relative Atom-Level Affinity Modeling

As shown in Fig. 1 (Top), competitive outcomes are strongly associated with the relative binding affinity between the test ligand and the radioligand. Intuitively, when the test ligand forms more

favorable interactions with the shared protein pocket than the radioligand, it is more likely to displace the radioligand and induce stronger inhibition, leading to a higher measured pIC_{50}. To encode this competition-aware signal, we further develop atom-level affinity features Thaingtamtanha et al. (2025) derived from AutoDock Vina scores Trott & Olson (2010) and computed using SMINA Koes et al. (2013).

To make this affinity signal directly accessible to the competitive graph, we use these Vina-derived descriptors as ligand atom-level node features. For each atom in the test ligand or radioligand, we compute a local protein–ligand affinity descriptor that summarizes its interaction pattern with the surrounding protein atoms. The resulting descriptor is concatenated with the original atom representation of the corresponding ligand node, thereby providing the model with fine-grained local affinity information for comparing the two competing ligands under the same protein environment.

Specifically, for each ligand atom, Smina decomposes the protein–ligand interaction into five empirical scoring terms, including two Gaussian steric terms, a steric repulsion term, a hydrophobic interaction term, and a non-directional hydrogen-bonding term. These terms correspond to the default Vina-style scoring components used in Smina, where the Gaussian, repulsion, hydrophobic, and hydrogen-bond terms are defined over pairwise atom interactions within a distance cutoff. For a ligand atom i and a protein atom j, let r_{ij} denote their Euclidean distance in Å, and let w_i and w_j denote their van der Waals radii. The effective surface distance is defined as

$$d_{ij} = r_{ij} - w_i - w_j. \tag{5}$$

Based on this surface distance, the pairwise Vina scoring terms are computed as

$$\mathrm{Gauss}_1(d_{ij}) = \exp\left[-\left(\frac{d_{ij}}{0.5}\right)^2\right], \tag{6}$$

$$\mathrm{Gauss}_2(d_{ij}) = \exp\left[-\left(\frac{d_{ij}-3.0}{2.0}\right)^2\right], \tag{7}$$

$$\mathrm{Repulsion}(d_{ij}) = \begin{cases} d_{ij}^2, & d_{ij} < 0, \\ 0, & d_{ij} \geq 0, \end{cases} \tag{8}$$

$$\mathrm{Hydrophobic}(d_{ij}) = \begin{cases} 1, & d_{ij} < 0.5, \\ 1.5 - d_{ij}, & 0.5 \leq d_{ij} \leq 1.5, \\ 0, & d_{ij} > 1.5, \end{cases} \tag{9}$$

and

$$\mathrm{HBond}(d_{ij}) = \begin{cases} 1, & d_{ij} < -0.7, \\ -\frac{d_{ij}}{0.7}, & -0.7 \leq d_{ij} \leq 0, \\ 0, & d_{ij} > 0. \end{cases} \tag{10}$$

Here, the Gaussian terms capture favorable steric complementarity at different distance ranges, the repulsion term penalizes steric clashes, and the hydrophobic and hydrogen-bonding terms describe chemically specific non-covalent interactions.

Since these scoring terms are defined at the atom-pair level, they are aggregated over the neighboring protein atoms to obtain a per-ligand-atom affinity descriptor. For a ligand atom i, its atom-level Vina feature is defined as

$$\begin{aligned} \mathbf{s}_i = \sum_{j \in \mathcal{N}(i)} \Big[& \mathrm{Gauss}_1(d_{ij}), \mathrm{Gauss}_2(d_{ij}), \\ & \mathrm{Repulsion}(d_{ij}), \mathrm{Hydrophobic}(d_{ij}), \mathrm{HBond}(d_{ij}) \Big] \end{aligned} \tag{11}$$

where $\mathcal{N}(i)$ denotes the set of neighboring protein atoms within the 8 Å scoring cutoff used by AutoDock Vina Trott & Olson (2010). In addition to preserving the individual interaction components, we further compute their weighted sum using the original Vina scoring weights to obtain an overall atom-level Vina score:

$$s_i^{\mathrm{Vina}} = \mathbf{w}^{\mathrm{T}} \mathbf{s}_i, \tag{12}$$

where $\mathbf{w}$ denotes the weights associated with the five Vina interaction terms Trott & Olson (2010). The final atom-level descriptor is therefore given by

$$\tilde{\mathbf{s}}_i = \left[\mathbf{s}_i,\, s_i^{\mathrm{Vina}}\right], \tag{13}$$

which produces a six-dimensional vector for each ligand atom, summarizing both its decomposed local interaction profile and its overall Vina-based affinity contribution within the surrounding binding pocket.

3.3 Steric Overlap Modeling via Inter-Ligand Edges

Fig. 1 (Bottom) highlights that competitive binding is not determined solely by the individual affinity of each ligand to the protein pocket, but is also shaped by the spatial relationship between their preferred binding poses. Since radioligands are typically designed to occupy the functional binding region within the pocket W. Pike (2016), test ligands that target the same region will be more likely to displace the radioligand, thereby substantially influencing the measured bioactivity. To explicitly encode this effect, we introduce inter-ligand edges between the test ligand and the radioligand.

Specifically, we extend the competitive triplet graph by adding an inter-ligand edge set $\mathcal{E}_{lr}$. An edge $e_{ij} \in \mathcal{E}_{lr}$ is constructed between atom $v_i \in \mathcal{V}_l$ from the test ligand and atom $v_j \in \mathcal{V}_r$ from the radioligand when their Euclidean distance is smaller than 6 Å:

$$e_{ij} \in \mathcal{E}_{lr} \quad \text{if} \quad d_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|_2 < 6 \text{ Å}. \tag{14}$$

These inter-ligand edges connect spatially proximal atoms across the two ligands, allowing the graph to capture local pose overlap within the shared binding pocket.

The resulting edge set enables direct message passing between competing ligand nodes, as illustrated by the dotted arrows in Fig. 3, providing a more explicit representation of ligand competition than relying only on ligand–protein interactions. Without $\mathcal{E}_{lr}$, information between the test ligand and radioligand must propagate indirectly through protein nodes, which may be weakened by the limited depth of graph neural networks due to over-smoothing Chen et al. (2020). By introducing $\mathcal{E}_{lr}$, the model can directly exchange information between spatially overlapping ligand regions, distinguishing cases where the two ligands occupy similar pocket volumes from those where their preferred poses are separated.

4 Experiments

4.1 Experimental Settings

Baselines. To evaluate the generalizability of DrugComp, we applied the framework to seven representative bioactivity prediction architectures spanning both transformer-based and graph-based models: transformer-based HypSeek Wang et al. (2026) and T-ALPHA Kyro et al. (2025), and graph-based GEMS Graber et al. (2025), CheapNet Lim et al. (2025), DTIGN Yin et al. (2024), GIGN Yang et al. (2023), and EGNN Satorras et al. (2021). DrugComp augments each baseline by explicitly modeling ligand competition through local binding scores and competitive graph representations.

Metrics. We report RMSE, MAE, Pearson (PR), Spearman (SR) and Kendall rank (TAU) correlations, and Precision@k, computed against compounds experimentally active at $\mathrm{IC}_{50} \leq 100$ nM Deng et al. (2022); Landrum & Riniker (2024). Avg Imp is the average relative improvement over these metrics against the corresponding baseline. All results use five-fold cross-validation; formal definitions are given in Appendix E.

4.2 Experimental Results and Analysis

Performance on CompDB-GPCR and CompDB-SLC Datasets. As shown in Table 1, DrugComp consistently improves all seven backbones on both benchmarks and across all reported metrics. On CompDB-GPCR (Top), average gains range from 2.18% to 11.48%: RMSE and MAE are reduced for every model, most on HypSeek (9.03% and 10.20%) and T-ALPHA (8.09% and 9.23%), while the largest correlation gains in PR, SR and Tau are observed for HypSeek (15.45%, 15.18%,

Table 1: Benchmark Performance on the CompDB-GPCR and CompDB-SLC Dataset. Results are shown as mean ± standard deviation over 5-fold cross-validation. The best values for each model are highlighted in bold.

Dataset	Model	Variant	RMSE (↓)	MAE (↓)	PR (↑)	SR (↑)	Tau (↑)	Prec@100 (↑)	Avg Imp (↑)
CompDB -GPCR	HypSeek (2026)	Baseline	1.039±0.043	0.827±0.037	0.579±0.042	0.574±0.040	0.406±0.032	0.958±0.027	-
		DrugComp	**0.945±0.029**	**0.743±0.023**	**0.668±0.017**	**0.662±0.012**	**0.480±0.011**	**0.964±0.013**	**11.48%**
	T-ALPHA (2025)	Baseline	0.981±0.015	0.765±0.019	0.637±0.026	0.631±0.026	0.453±0.023	0.942±0.019	-
		DrugComp	**0.902±0.024**	**0.694±0.014**	**0.706±0.015**	**0.699±0.017**	**0.515±0.012**	**0.976±0.028**	**9.36%**
	GEMS (2025)	Baseline	0.857±0.008	0.652±0.008	0.738±0.008	0.726±0.006	0.544±0.007	0.970±0.023	-
		DrugComp	**0.795±0.018**	**0.598±0.012**	**0.781±0.009**	**0.772±0.008**	**0.588±0.008**	**0.990±0.000**	**6.32%**
	CheapNet (2025)	Baseline	0.891±0.019	0.682±0.017	0.715±0.014	0.703±0.014	0.521±0.012	0.968±0.019	-
		DrugComp	**0.869±0.026**	**0.664±0.019**	**0.728±0.015**	**0.720±0.012**	**0.536±0.011**	**0.978±0.013**	**2.18%**
	DTIGN (2024)	Baseline	0.895±0.044	0.694±0.038	0.706±0.033	0.698±0.033	0.514±0.029	0.966±0.018	-
		DrugComp	**0.861±0.027**	**0.663±0.025**	**0.735±0.017**	**0.726±0.015**	**0.541±0.013**	**0.980±0.012**	**3.82%**
	GIGN (2023)	Baseline	0.904±0.034	0.690±0.023	0.707±0.018	0.699±0.024	0.516±0.020	0.962±0.026	-
		DrugComp	**0.869±0.025**	**0.663±0.017**	**0.731±0.013**	**0.722±0.014**	**0.537±0.012**	**0.978±0.019**	**3.36%**
	EGNN (2021)	Baseline	0.974±0.040	0.766±0.036	0.640±0.033	0.630±0.034	0.453±0.029	0.964±0.021	-
		DrugComp	**0.901±0.027**	**0.703±0.023**	**0.703±0.021**	**0.696±0.020**	**0.510±0.018**	**0.972±0.023**	**8.27%**
CompDB -SLC	HypSeek (2026)	Baseline	1.012±0.118	0.790±0.088	0.485±0.133	0.516±0.126	0.355±0.094	0.860±0.114	-
		DrugComp	**0.893±0.074**	**0.694±0.061**	**0.614±0.050**	**0.624±0.060**	**0.439±0.045**	**0.920±0.084**	**17.05%**
	T-ALPHA (2025)	Baseline	1.013±0.066	0.802±0.063	0.492±0.126	0.497±0.115	0.342±0.081	0.620±0.217	-
		DrugComp	**0.918±0.105**	**0.710±0.068**	**0.620±0.085**	**0.629±0.077**	**0.446±0.060**	**0.860±0.114**	**23.77%**
	GEMS (2025)	Baseline	0.930±0.087	0.720±0.046	0.581±0.092	0.591±0.076	0.421±0.054	0.820±0.130	-
		DrugComp	**0.898±0.034**	**0.692±0.021**	**0.627±0.031**	**0.644±0.023**	**0.459±0.022**	**0.840±0.182**	**6.00%**
	CheapNet (2025)	Baseline	0.894±0.079	0.677±0.058	0.655±0.057	0.664±0.055	0.487±0.046	0.840±0.114	-
		DrugComp	**0.822±0.041**	**0.626±0.016**	**0.708±0.024**	**0.718±0.019**	**0.525±0.012**	**0.880±0.084**	**7.39%**
	DTIGN (2024)	Baseline	0.980±0.065	0.780±0.044	0.564±0.052	0.575±0.056	0.406±0.043	0.800±0.187	-
		DrugComp	**0.900±0.076**	**0.695±0.041**	**0.643±0.055**	**0.658±0.052**	**0.471±0.038**	**0.920±0.084**	**13.08%**
	GIGN (2023)	Baseline	0.919±0.050	0.723±0.036	0.618±0.047	0.630±0.059	0.449±0.046	0.800±0.100	-
		DrugComp	**0.874±0.075**	**0.687±0.063**	**0.669±0.048**	**0.680±0.048**	**0.492±0.043**	**0.920±0.084**	**8.46%**
	EGNN (2021)	Baseline	1.041±0.086	0.830±0.072	0.453±0.096	0.454±0.131	0.308±0.094	0.740±0.167	-
		DrugComp	**0.904±0.048**	**0.695±0.037**	**0.623±0.052**	**0.634±0.041**	**0.446±0.033**	**0.820±0.045**	**26.99%**

Table 2: Ablation Study Performance on the CompDB-AR Dataset. Results are shown as mean ± standard deviation over 5-fold cross-validation. The best and second-best results for each model are highlighted in bold and underlined, respectively.

Model	Variant	RMSE (↓)	MAE (↓)	PR (↑)	SR (↑)	Tau (↑)	Prec@10 (↑)	Avg Imp (↑)
HypSeek (2026)	Baseline	0.884 ± 0.057	0.718 ± 0.073	0.711 ± 0.064	0.659 ± 0.078	0.494 ± 0.066	0.780 ± 0.164	-
	+CGR	0.890 ± 0.066	0.720 ± 0.071	0.696 ± 0.075	0.675 ± 0.122	0.506 ± 0.103	0.840 ± 0.152	1.60%
	+CGR-LBS	$\underline{0.879 \pm 0.154}$	$\underline{0.707 \pm 0.131}$	0.694 ± 0.132	0.678 ± 0.146	0.509 ± 0.133	$\underline{0.860 \pm 0.114}$	2.66%
	+CGR-ILE	0.895 ± 0.113	0.735 ± 0.099	$\underline{0.717 \pm 0.057}$	$\underline{0.700 \pm 0.096}$	$\underline{0.529 \pm 0.080}$	$\mathbf{0.880 \pm 0.045}$	$\underline{3.92\%}$
	Full	$\mathbf{0.806 \pm 0.036}$	$\mathbf{0.640 \pm 0.033}$	$\mathbf{0.753 \pm 0.039}$	$\mathbf{0.749 \pm 0.086}$	$\mathbf{0.574 \pm 0.083}$	0.860 ± 0.114	$\mathbf{11.00\%}$
T-ALPHA (2025)	Baseline	1.095 ± 0.237	0.874 ± 0.256	0.475 ± 0.228	0.452 ± 0.245	0.338 ± 0.197	0.760 ± 0.182	-
	+CGR	1.044 ± 0.350	0.828 ± 0.311	0.570 ± 0.228	0.532 ± 0.281	0.397 ± 0.217	0.780 ± 0.179	11.28%
	+CGR-LBS	1.012 ± 0.179	0.820 ± 0.152	0.568 ± 0.193	0.528 ± 0.243	0.390 ± 0.201	$\underline{0.840 \pm 0.152}$	12.69%
	+CGR-ILE	$\underline{0.970 \pm 0.089}$	$\underline{0.781 \pm 0.099}$	$\underline{0.616 \pm 0.099}$	$\underline{0.580 \pm 0.130}$	$\underline{0.428 \pm 0.119}$	0.760 ± 0.114	$\underline{17.76\%}$
	Full	$\mathbf{0.954 \pm 0.083}$	$\mathbf{0.761 \pm 0.090}$	$\mathbf{0.650 \pm 0.099}$	$\mathbf{0.627 \pm 0.089}$	$\mathbf{0.463 \pm 0.072}$	$\mathbf{0.840 \pm 0.089}$	$\mathbf{24.78\%}$
GEMS (2025)	Baseline	1.014 ± 0.094	0.818 ± 0.113	0.586 ± 0.110	0.523 ± 0.148	0.389 ± 0.119	0.740 ± 0.134	-
	+CGR	0.903 ± 0.151	0.738 ± 0.155	0.705 ± 0.099	0.669 ± 0.114	0.508 ± 0.101	0.780 ± 0.148	17.52%
	+CGR-LBS	$\underline{0.849 \pm 0.074}$	$\underline{0.677 \pm 0.076}$	$\underline{0.727 \pm 0.049}$	$\mathbf{0.697 \pm 0.041}$	$\mathbf{0.536 \pm 0.046}$	0.760 ± 0.089	$\underline{21.91\%}$
	+CGR-ILE	0.875 ± 0.142	0.698 ± 0.123	0.694 ± 0.108	0.653 ± 0.114	0.497 ± 0.096	$\underline{0.840 \pm 0.089}$	18.81%
	Full	$\mathbf{0.824 \pm 0.109}$	$\mathbf{0.647 \pm 0.090}$	$\mathbf{0.734 \pm 0.073}$	$\underline{0.685 \pm 0.059}$	$\underline{0.520 \pm 0.049}$	$\mathbf{0.880 \pm 0.084}$	$\mathbf{24.73\%}$

18.40%), followed by T-ALPHA (10.88%, 10.72%, 13.61%) and EGNN (9.87%, 10.44%, 12.71%); Precision@100 increases for every backbone, led by T-ALPHA (+3.61%), GEMS (+2.06%) and GIGN (+1.66%). On CompDB-SLC (Bottom) the gains are larger, averaging 26.99% for EGNN, 23.77% for T-ALPHA and 17.05% for HypSeek; EGNN leads both in error reduction (13.19% RMSE, 16.28% MAE) and in correlation (37.34% PR, 39.60% SR, 44.73% Tau), with HypSeek, T-ALPHA and DTIGN also improving substantially, and Precision@10 improves for all models, particularly T-ALPHA, DTIGN and GIGN. Overall, DrugComp provides robust and architecture-agnostic improvements in prediction accuracy, ranking consistency, and early hit retrieval.

Ablation Study on CompDB-AR Dataset. To quantify the contribution of each component in DrugComp, we perform ablation studies on the CompDB-AR dataset, as shown in Table 2. The complete results for all seven evaluated architectures reported in Appendix Table 5. Across all evaluated architectures, incorporating competitive graph representation (+CGR) improves over the

Table 3: Effect of the competitor on CompDB-GPCR. Each model is evaluated with no competitor, with the assayed radioligand, with a different radioligand of the same target, and with an arbitrary test ligand of the same target. $\Delta\%$ is the increase in test RMSE incurred by each alternative relative to the assayed radioligand; p is the p-value from paired t-tests over the five folds.

	test RMSE by competitor				vs. no competitor		vs. other radioligand		vs. decoy ligand	
Backbone	none	**assayed**	other radioligand	decoy	$\Delta\%$	p	$\Delta\%$	p	$\Delta\%$	p
HypSeek	1.060	**0.998**	1.033	1.074	+5.9%	0.012	+3.5%	0.049	+7.1%	0.000
GEMS	0.915	**0.860**	0.918	0.936	+6.0%	0.018	+6.3%	0.010	+8.1%	0.005
DTIGN	0.920	**0.903**	0.948	0.979	+1.8%	0.201	+4.7%	0.001	+7.8%	0.005

Table 4: Test RMSE under group-disjoint splits of CompDB-GPCR. Mean $\pm$ standard deviation over five folds. Bold marks a DrugComp value better than its baseline.

	Ligand-disjoint		Scaffold-disjoint		Target-disjoint	
Backbone	Baseline	DrugComp	Baseline	DrugComp	Baseline	DrugComp
HypSeek	1.031±0.020	**0.947**±0.027	1.094±0.061	**1.053**±0.025	1.322±0.146	**1.267**±0.100
GEMS	0.854±0.014	**0.849**±0.032	0.997±0.035	**0.985**±0.061	1.271±0.150	**1.257**±0.111
DTIGN	0.888±0.014	**0.878**±0.021	0.993±0.021	**0.967**±0.044	1.419±0.183	**1.388**±0.196

baseline, confirming the benefit of jointly modeling the pocket, test ligand, and radioligand. The addition of local binding scores (+CGR-LBS) further improves performance over +CGR across most models, indicating that atom-level energetic cues provide complementary information to structural representations. Similarly, introducing inter-ligand edges (+CGR-ILE) provides additional gains on top of +CGR, suggesting that explicitly modeling inter-ligand spatial overlaps is beneficial. Overall, the full model achieves the best performance across all evaluated architectures, demonstrating that these components jointly provide robust and generalizable improvements.

Which competitor the gain depends on. Does any competitor help, or only the one the assay used? Table 3 reports a separate experiment on the same $1{,}266$ assays per fold. The test ligand and its label are fixed throughout; only the competitor in the pocket is exchanged. Removing the competitor raises test RMSE by 1.8–6.0% over the assayed radioligand. Putting a different real radioligand of the target there raises it by 3.5–6.3% ($p \leq 0.049$), and putting an arbitrary test ligand there, a real molecule never used as a radioligand, by 7.1–8.1% ($p \leq 0.005$). Only the assayed radioligand helps, which points to the underlying biology rather than to the presence of a second molecule.

What models draw from the competitor. Two experiments ask whether the gain is an artifact of the measurement or of the split. (*i*) *A competitor-independent label.* One worry is that the model only recovers the radioligand's concentration or affinity, since IC_{50} depends on them through the Cheng-Prusoff relation $K_i = \mathrm{IC}_{50}/(1 + [L]/K_d)$, with $[L]$ the radioligand concentration and K_d its affinity. A K_i label has that factor divided out, so it carries no information about $[L]$ or K_d for the model to recover—yet DrugComp still improves over its competition-blind baseline, retaining 89% of the improvement with GEMS and 69% with HypSeek (Appendix B). (*ii*) *Beyond random splits.* Replacing random splits with ligand-, scaffold- and target-disjoint splits, DrugComp lowers test RMSE in all nine backbone–split comparisons (p=0.0039; Table 4), improving every backbone by 1.1–4.2% on unseen targets.

5 Conclusion

In this work, we revisit bioactivity prediction from the perspective of ligand competition, a key factor that is largely overlooked in existing methods. We present DrugComp, a competition-aware framework for structure-based bioactivity prediction that models ligand competition within shared protein binding pockets. By integrating local binding scores and inter-ligand edges into unified competitive binding graphs, DrugComp captures both atom-level affinity cues and spatial ligand-overlap patterns, leading to improved prediction accuracy and ranking performance across diverse architectures and protein families. Alongside DrugComp, we introduce CompDB, a curated dataset that links test ligands with experimentally characterized radioligands in shared binding environments, providing a benchmark for competition-aware learning. Overall, our work highlights the importance of incorporating ligand competition into bioactivity prediction and lays the foundation for more realistic and context-aware drug discovery models.

AI USE STATEMENT

Generative AI tools were used in this work in a limited and supporting capacity. As part of the construction of CompDB, we used Meta-Llama-3.8B-Instruct to extract radioligand identities from unstructured radioligand displacement assay descriptions. This use was restricted to structured information extraction from existing assay metadata and did not involve generating experimental observations, bioactivity labels, molecular structures, or synthetic data. The extracted radioligand identities were subsequently mapped to canonical molecular representations using the ChEML/ChEMBL data resources described in the dataset construction procedure. To verify the reliability of this AI-assisted extraction step, we manually reviewed a randomly sampled subset of 100 assays and confirmed the correctness of the extracted radioligand identities.

Generative AI tools were additionally used to edit and polish portions of the manuscript for grammar, clarity, and readability, without changing the intended scientific content, and to assist with modifications to existing software code, such as code editing, debugging, refactoring, and resolving implementation issues. AI-assisted code changes were reviewed and tested by the authors before being incorporated into the experiments. Generative AI tools did not determine the algorithmic design of the proposed method or independently implement the proposed methodology.

Except for the radioligand information-extraction step described above, we did not use generative AI tools to generate synthetic datasets; develop theoretical models or conceptual frameworks; formulate mathematical claims; provide ingredients for mathematical proofs; write proofs; propose or refine research hypotheses; design or provide feedback on the research methodology or experiments; implement the proposed methods; assist with translation; clean or reformat datasets; perform qualitative or thematic data analysis; or interpret experimental results. We also did not rely on generative AI tools to determine the scientific conclusions reported in this paper.

All AI-assisted text, code, and extracted information used in the final work were reviewed by the authors. The authors independently verified the scientific content, experimental results, and conclusions, and take full responsibility for the final manuscript and all claims and artifacts contained therein.

REPRODUCIBILITY STATEMENT

We have made substantial efforts to provide the information necessary to reproduce the experiments reported in this work. The construction of CompDB, including assay selection, radioligand extraction and validation, binding-site identification, docking, dataset organization, and graph construction, is described in the *CompDB Dataset Construction* section, with additional dataset statistics provided in Appendix D. The complete formulation of DrugComp, including the competitive binding graph, atom-level affinity features, and inter-ligand edges, is provided in Section 3. Definitions of all evaluation metrics are given in Appendix E, while Appendix F documents the evaluated backbone models, model-specific graph representations and features, hyperparameter settings and search ranges, dataset splitting procedure, random seeds, software environment, optimization settings, early-stopping criterion, and computational resources used for training. Together, these sections specify the data-processing, structural-modeling, training, and evaluation procedures used to obtain the reported results.

The source code is not released because the proposed methodology is subject to ongoing intellectual-property protection. This restriction concerns code availability rather than disclosure of the methodology; the algorithmic formulation and implementation choices required to reconstruct the approach are described in the main paper and appendix as detailed above. The CompDB dataset will be made publicly available upon publication of the paper to facilitate independent evaluation and future research. The reported results are based on five-fold cross-validation with fixed dataset-splitting procedures, and the corresponding implementation and experimental details are provided in the appendix to support independent reproduction of the study.

REFERENCES

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

Dominique Bonafoux, Ayome Abibi, Brian Bettencourt, Andrew Burchat, Anna Ericsson, Christopher M Harris, Tegest Kebede, Michael Morytko, Michael McPherson, Grier Wallace, et al. Thienopyrrole acetic acids as antagonists of the crth2 receptor. *Bioorganic & medicinal chemistry letters*, 21(6):1861–1864, 2011.

Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou, and Xu Sun. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 3438–3445, 2020.

UniProt Consortium. Uniprot: a hub for protein information. *Nucleic acids research*, 43(D1): D204–D212, 2015.

Dingsheng Deng. Dbscan clustering algorithm based on density. In *2020 7th international forum on electrical engineering and automation (IFEEA)*, pp. 949–953. IEEE, 2020.

Jianyuan Deng, Zhibo Yang, Iwao Ojima, Dimitris Samaras, and Fusheng Wang. Artificial intelligence in drug discovery: applications and techniques. *Briefings in Bioinformatics*, 23(1): bbab430, 2022.

Jakub Fichna, Renata Perlikowska, Anna Wyrebska, Katarzyna Gach, Justyna Piekielna, Jean Claude do Rego, Geza Toth, Alicja Kluczyk, Tomasz Janecki, and Anna Janecka. Effect of 2', 6'-dimethyl-l-tyrosine (dmt) on pharmacological activity of cyclic endomorphin-2 and morphiceptin analogs. *Bioorganic & medicinal chemistry*, 19(23):6977–6981, 2011.

Anna Gaulton, Louisa J Bellis, A Patricia Bento, Jon Chambers, Mark Davies, Anne Hersey, Yvonne Light, Shaun McGlinchey, David Michalovich, Bissan Al-Lazikani, et al. Chembl: a large-scale bioactivity database for drug discovery. *Nucleic acids research*, 40(D1):D1100–D1107, 2012.

David Graber, Peter Stockinger, Fabian Meyer, Siddhartha Mishra, Claus Horn, and Rebecca Buller. Resolving data bias improves generalization in binding affinity prediction. *Nature Machine Intelligence*, pp. 1–13, 2025.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Bowei He, Bowen Gao, Yankai Chen, Yanyan Lan, Chen Ma, Philip S Yu, Ya-Qin Zhang, and Wei-Ying Ma. S^2drug: Bridging protein sequence and 3d structure in contrastive representation learning for virtual screening. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 354–362, 2026.

Yanwen Huang, Bowen Gao, Yinjun Jia, Hongbo Ma, Wei-Ying Ma, Ya-Qin Zhang, and Yanyan Lan. Redefining the task of bioactivity prediction. In *ICLR*, 2025.

David Ryan Koes, Matthew P Baumgartner, and Carlos J Camacho. Lessons learned in empirical scoring with smina from the csar 2011 benchmarking exercise. *Journal of chemical information and modeling*, 53(8):1893–1904, 2013.

Gregory W Kyro, Anthony M Smaldone, Yu Shee, Chuzhi Xu, and Victor S Batista. T-alpha: A hierarchical transformer-based deep neural network for protein–ligand binding affinity prediction with uncertainty-aware self-learning for protein-specific alignment. *Journal of Chemical Information and Modeling*, 65(5):2395–2415, 2025.

Gregory A Landrum and Sereina Riniker. Combining ic50 or k i values from different sources is a source of significant noise. *Journal of chemical information and modeling*, 64(5):1560–1567, 2024.

Jiashan Li and Xinqi Gong. Harnessing pre-trained models for accurate prediction of protein-ligand binding affinity. *BMC bioinformatics*, 26(1):1–21, 2025.

Hyukjun Lim, Sun Kim, and Sangseon Lee. Cheapnet: Cross-attention on hierarchical representations for efficient protein-ligand binding affinity prediction. In *ICLR*, 2025.

Janet J. Maguire, Rhoda E. Kuc, and Anthony P. Davenport. Radioligand binding assays and their analysis. In *Receptor Binding Techniques*, pp. 31–77. Springer, 2012.

Erik Martinez-Hackert, Anders Sundan, and Toril Holien. Receptor binding competition: A paradigm for regulating tgf-β family action. *Cytokine & growth factor reviews*, 57:39–54, 2021.

Andrea Mastropietro, Giuseppe Pasculli, and Jürgen Bajorath. Learning characteristics of graph neural networks predicting protein–ligand affinities. *Nature Machine Intelligence*, pp. 1427–1436, 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 Özkirimli Ölmez. Deepdta: deep drug-target binding affinity prediction. *Bioinform.*, 2018.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv*, 2025. doi: 10.1101/2025.06.14.659707.

Peter W Rose, Andreas Prlić, Ali Altunkaya, Chunxiao Bi, Anthony R Bradley, Cole H Christie, Luigi Di Costanzo, Jose M Duarte, Shuchismita Dutta, Zukang Feng, et al. The rcsb protein data bank: integrative view of protein, gene and 3d structural information. *Nucleic acids research*, pp. gkw1000, 2016.

Anastasiia V Sadybekov and Vsevolod Katritch. Computational approaches streamlining drug discovery. *Nature*, 616(7958):673–685, 2023.

Severo Salvadori, Mauro Marastoni, Gianfranco Balboni, Pier Andrea Borea, Michele Morari, and Roberto Tomatis. Synthesis and structure-activity relationships of deltorphins analogs. *Journal of medicinal chemistry*, 34(5):1656–1661, 1991.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Linde Schoenmaker, Enzo G Sastrokarijo, Laura H Heitman, Joost B Beltman, Willem Jespers, and Gerard JP van Westen. Toward assay-aware bioactivity model (er) s: Getting a grip on biological context. *Journal of Chemical Information and Modeling*, 2025.

Bonggun Shin, Sungsoo Park, Keunsoo Kang, and Joyce C Ho. Self-attention based molecule representation for predicting drug-target interaction. In *Machine learning for healthcare conference*. PMLR, 2019.

Thanawat Thaingtamtanha, Jordane Preto, and Francesco Gentile. Pose ensemble graph neural networks to improve docking performances. *Chemical Science*, 16(42):19876–19887, 2025.

Oleg Trott and Arthur J Olson. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.

Victor W. Pike. Considerations in the development of reversibly binding pet radioligands for brain imaging. *Current Medicinal Chemistry*, 23(18):1818–1869, 2016.

Guishen Wang, Hangchen Zhang, Mengting Shao, Shisen Sun, and Chen Cao. Mmpd-dta: Integrating multi-modal deep learning with pocket-drug graphs for drug-target binding affinity prediction. *Journal of Chemical Information and Modeling*, 65(3):1615–1630, 2025a.

Jianhui Wang, Wenyu Zhu, Bowen Gao, Xin Hong, Ya-Qin Zhang, Wei-Ying Ma, and Yanyan Lan. Learning protein–ligand binding in hyperbolic space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 1150–1158, 2026.

Kai Wang, Aijie Song, Fei Liu, Xiaoli Luan, Xinglong Wang, and Jingwen Zhou. Plmam-pla: A method using pretrained language models and attention mechanisms for protein-ligand binding affinity prediction. *IEEE Transactions on Computational Biology and Bioinformatics*, 2025b.

Zechen Wang, Sheng Wang, Yangyang Li, Jingjing Guo, Yanjie Wei, Yuguang Mu, Liangzhen Zheng, and Weifeng Li. A new paradigm for applying deep learning to protein–ligand interaction prediction. *Briefings in bioinformatics*, pp. bbae145, 2024.

Qianyang Wu, Jingwei Lv, Zilong Zhang, and Feifei Cui. Generalizable drug–target interaction prediction via esm-2 representations and progressive contrastive curriculum learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 1276–1284, 2026.

Yiming Yang, Zhiyuan Zhou, Yueming Yin, Hoi-Yeung Li, and Adams Wai-Kin Kong. Test-time adaptation without source data for out-of-domain bioactivity prediction. In *ICLR*, 2026.

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*, 2023.

Yueming Yin, Haifeng Hu, Zhen Yang, Huajian Xu, and Jiansheng Wu. Realvs: toward enhancing the precision of top hits in ligand-based virtual screening of drug leads from large compound databases. *Journal of chemical information and modeling*, 61(10):4924–4939, 2021.

Yueming Yin, Haifeng Hu, Zhen Yang, Feihu Jiang, Yihe Huang, and Jiansheng Wu. Afse: towards improving model generalization of deep graph learning of ligand bioactivities targeting gpcr proteins. *Briefings in Bioinformatics*, 23(3), 2022.

Yueming Yin, Hilbert Yuen In Lam, Yuguang Mu, Hoi-Yeung Li, and Adams Wai-Kin Kong. Advancing bioactivity prediction through molecular docking and self-attention. *IEEE J. Biomed. Health Informatics*, pp. 7599–7610, 2024.

Zhiyuan Zhou, Yueming Yin, Yiming Yang, Yuguang Mu, Hoi-Yeung Li, and Adams Wai Kin Kong. Enhancing bioactivity prediction via spatial emptiness representation of protein-ligand complex and union of multiple pockets. In *NeurIPS*, 2025.

A Complete Ablation Results on All Evaluated Architectures

To provide a more comprehensive analysis of the contribution of each component in DrugComp, we report the complete ablation results on all seven evaluated architectures in Table 5. The results show that the proposed components bring consistent improvements across diverse backbone models, including HypSeek, T-ALPHA, GEMS, CheapNet, DTIGN, GIGN, and EGNN. First, adding the competitive graph representation (+CGR) generally improves performance over the corresponding baseline. This demonstrates that explicitly constructing a joint graph over the binding pocket, test ligand, and radioligand provides useful competitive-context information that is missing from conventional ligand–protein modeling. Second, incorporating local binding scores (+CGR-LBS) further improves performance for most architectures, suggesting that atom-level energetic interactions provide complementary signals beyond geometric and topological representations. Third, adding inter-ligand edges (+CGR-ILE) also yields additional gains over +CGR, indicating that directly modeling the spatial relationship between the test ligand and radioligand helps capture competitive displacement effects. Finally, the full DrugComp model, which combines competitive graph representation, local binding scores, and inter-ligand edges, achieves the strongest overall performance across the seven evaluated architectures.

B The K_i Control

A K_i has the Cheng-Prusoff factor $(1 + [L]/K_d)$ divided out already, so it is corrected for the radioligand's concentration and affinity; a purely competitive contribution should therefore vanish under that label. We run a separate five-fold cross-validation on CompDB-GPCR for two representative backbones, once with an IC_{50} label and once with a K_i label, on matched sets of

Table 5: Ablation Study Performance on the CompDB-AR Dataset. Results are shown as mean ± standard deviation over 5-fold cross-validation. The best and second-best results for each model are highlighted in bold and underlined, respectively.

Model	Variant	RMSE (↓)	MAE (↓)	PR (↑)	SR (↑)	Tau (↑)	Prec@10 (↑)	Avg Imp (↑)
HypSeek (2026)	Baseline	0.884 ± 0.057	0.718 ± 0.073	0.711 ± 0.064	0.659 ± 0.078	0.494 ± 0.066	0.780 ± 0.164	-
	+CGR	0.890 ± 0.066	0.720 ± 0.071	0.696 ± 0.075	0.675 ± 0.122	0.506 ± 0.103	0.840 ± 0.152	1.60%
	+CGR-LBS	0.879 ± 0.154	0.707 ± 0.131	0.694 ± 0.132	0.678 ± 0.146	0.509 ± 0.133	0.860 ± 0.114	2.66%
	+CGR-ILE	0.895 ± 0.113	0.735 ± 0.099	0.717 ± 0.057	0.700 ± 0.096	0.529 ± 0.080	$\mathbf{0.880 \pm 0.045}$	3.92%
	Full	$\mathbf{0.806 \pm 0.036}$	$\mathbf{0.640 \pm 0.033}$	$\mathbf{0.753 \pm 0.039}$	$\mathbf{0.749 \pm 0.086}$	$\mathbf{0.574 \pm 0.083}$	0.860 ± 0.114	**11.00%**
T-ALPHA (2025)	Baseline	1.095 ± 0.237	0.874 ± 0.256	0.475 ± 0.228	0.452 ± 0.245	0.338 ± 0.197	0.760 ± 0.182	-
	+CGR	1.044 ± 0.350	0.828 ± 0.311	0.570 ± 0.228	0.532 ± 0.281	0.397 ± 0.217	0.780 ± 0.179	11.28%
	+CGR-LBS	1.012 ± 0.179	0.820 ± 0.152	0.568 ± 0.193	0.528 ± 0.243	0.390 ± 0.201	0.840 ± 0.152	12.69%
	+CGR-ILE	0.970 ± 0.089	0.781 ± 0.099	0.616 ± 0.099	0.580 ± 0.130	0.428 ± 0.119	0.760 ± 0.114	17.76%
	Full	$\mathbf{0.954 \pm 0.083}$	$\mathbf{0.761 \pm 0.090}$	$\mathbf{0.650 \pm 0.099}$	$\mathbf{0.627 \pm 0.089}$	$\mathbf{0.463 \pm 0.072}$	$\mathbf{0.840 \pm 0.089}$	**24.78%**
GEMS (2025)	Baseline	1.014 ± 0.094	0.818 ± 0.113	0.586 ± 0.110	0.523 ± 0.148	0.389 ± 0.119	0.740 ± 0.134	-
	+CGR	0.903 ± 0.151	0.738 ± 0.155	0.705 ± 0.099	0.669 ± 0.114	0.508 ± 0.101	0.780 ± 0.148	17.52%
	+CGR-LBS	0.849 ± 0.074	0.677 ± 0.076	0.727 ± 0.049	$\mathbf{0.697 \pm 0.041}$	$\mathbf{0.536 \pm 0.046}$	0.760 ± 0.089	21.91%
	+CGR-ILE	0.875 ± 0.142	0.698 ± 0.123	0.694 ± 0.108	0.653 ± 0.114	0.497 ± 0.096	0.840 ± 0.089	18.81%
	Full	$\mathbf{0.824 \pm 0.109}$	$\mathbf{0.647 \pm 0.090}$	$\mathbf{0.734 \pm 0.073}$	0.685 ± 0.059	0.520 ± 0.049	$\mathbf{0.880 \pm 0.084}$	**24.73%**
CheapNet (2025)	Baseline	0.915 ± 0.059	0.706 ± 0.036	0.702 ± 0.084	0.668 ± 0.113	0.513 ± 0.102	0.780 ± 0.217	-
	+CGR	0.861 ± 0.031	0.692 ± 0.028	0.725 ± 0.069	0.683 ± 0.083	0.518 ± 0.084	0.840 ± 0.055	3.68%
	+CGR-LBS	$\mathbf{0.828 \pm 0.084}$	$\mathbf{0.651 \pm 0.046}$	0.746 ± 0.080	0.716 ± 0.073	$\mathbf{0.550 \pm 0.068}$	0.860 ± 0.114	8.02%
	+CGR-ILE	0.850 ± 0.124	0.684 ± 0.121	0.720 ± 0.067	0.689 ± 0.065	0.523 ± 0.070	0.880 ± 0.084	5.15%
	Full	0.841 ± 0.052	0.675 ± 0.058	$\mathbf{0.747 \pm 0.051}$	$\mathbf{0.717 \pm 0.060}$	0.539 ± 0.059	$\mathbf{0.920 \pm 0.045}$	**8.24%**
DTIGN (2024)	Baseline	0.936 ± 0.165	0.751 ± 0.148	0.680 ± 0.066	0.624 ± 0.073	0.468 ± 0.062	0.860 ± 0.152	-
	+CGR	0.894 ± 0.139	0.711 ± 0.116	0.705 ± 0.035	0.657 ± 0.061	0.500 ± 0.056	0.880 ± 0.084	4.65%
	+CGR-LBS	0.905 ± 0.093	0.727 ± 0.087	0.706 ± 0.070	$\mathbf{0.674 \pm 0.104}$	0.506 ± 0.105	0.860 ± 0.114	4.43%
	+CGR-ILE	0.868 ± 0.100	$\mathbf{0.662 \pm 0.061}$	0.697 ± 0.054	0.664 ± 0.137	0.507 ± 0.116	0.860 ± 0.089	6.07%
	Full	$\mathbf{0.867 \pm 0.118}$	0.663 ± 0.088	$\mathbf{0.711 \pm 0.076}$	0.669 ± 0.116	$\mathbf{0.508 \pm 0.103}$	$\mathbf{0.920 \pm 0.084}$	**7.70%**
GIGN (2023)	Baseline	0.912 ± 0.135	0.746 ± 0.123	0.669 ± 0.126	0.635 ± 0.141	0.470 ± 0.121	0.780 ± 0.179	-
	+CGR	0.843 ± 0.054	0.673 ± 0.058	0.718 ± 0.101	0.659 ± 0.140	0.503 ± 0.131	0.800 ± 0.187	6.32%
	+CGR-LBS	0.847 ± 0.110	0.670 ± 0.105	0.718 ± 0.094	0.693 ± 0.098	0.523 ± 0.083	$\mathbf{0.880 \pm 0.045}$	9.58%
	+CGR-ILE	0.873 ± 0.063	0.702 ± 0.076	0.705 ± 0.024	0.690 ± 0.055	0.526 ± 0.044	0.840 ± 0.089	7.26%
	Full	$\mathbf{0.818 \pm 0.104}$	$\mathbf{0.663 \pm 0.076}$	$\mathbf{0.762 \pm 0.076}$	$\mathbf{0.730 \pm 0.089}$	$\mathbf{0.553 \pm 0.094}$	0.820 ± 0.164	**12.18%**
EGNN (2021)	Baseline	1.023 ± 0.161	0.830 ± 0.155	0.549 ± 0.173	0.512 ± 0.216	0.383 ± 0.177	0.740 ± 0.152	-
	+CGR	1.028 ± 0.135	0.826 ± 0.138	0.600 ± 0.192	0.598 ± 0.192	0.451 ± 0.158	0.820 ± 0.179	9.14%
	+CGR-LBS	$\mathbf{0.949 \pm 0.121}$	$\mathbf{0.754 \pm 0.118}$	0.626 ± 0.140	0.613 ± 0.118	0.457 ± 0.120	0.800 ± 0.100	12.90%
	+CGR-ILE	0.958 ± 0.219	0.763 ± 0.206	0.612 ± 0.162	0.586 ± 0.135	0.432 ± 0.117	0.800 ± 0.158	10.24%
	Full	0.986 ± 0.172	0.788 ± 0.157	$\mathbf{0.631 \pm 0.117}$	$\mathbf{0.636 \pm 0.115}$	$\mathbf{0.472 \pm 0.100}$	$\mathbf{0.840 \pm 0.152}$	**14.10%**

∼2,030 test rows per fold so that training-set size is not a confound. The two label sets share only 2.63% of assay identifiers, so these are largely independent measurements rather than converted values. The gain nonetheless persists: GEMS retains 89% of its improvement and HypSeek 69%, both significant (Table 6).

Table 6: K_i control on CompDB-GPCR. Blind RMSE is the test RMSE of the competition-blind baselin; Gain is that value minus DrugComp's test RMSE, with p from paired t-tests over the five folds; Kept is the K_i gain as a fraction of the IC_{50} gain.

	IC_{50}			K_i			
Backbone	Blind RMSE	Gain	p	Blind RMSE	Gain	p	Kept
GEMS	0.862	+0.045	0.003	1.077	+0.040	0.020	89%
HypSeek	1.017	+0.074	0.014	1.184	+0.051	0.009	69%

C Related Work

Early deep learning (DL)-based approaches primarily focused on ligand-based representations, modeling small molecules as two-dimensional graphs (Yin et al., 2021; 2022). Subsequent methods incorporated protein sequence information to capture protein–ligand interactions. For example, DeepDTA (Öztürk et al., 2018) employed convolutional neural networks (CNNs) to learn latent representations from protein sequences and ligand SMILES, while MT-DTI (Shin et al., 2019) introduced attention mechanisms to capture long-range dependencies.

With the increasing availability of structural data, graph neural networks (GNNs) have emerged as a natural framework for modeling protein–ligand systems by representing atoms as nodes and

their interactions as edges. GraphDTA (Nguyen et al., 2021) was among the first to introduce molecular graphs for drug representation in this context. Subsequent studies further advanced fully structure-based modeling: methods such as GIGN (Yang et al., 2023) and DTIGN (Yin et al., 2024) constructed pocket–ligand complex graphs and performed end-to-end learning on protein–ligand interaction structures, enabling more fine-grained modeling of binding (Mastropietro et al., 2023; Wang et al., 2024). As a result, modeling protein pockets and ligand binding geometries has become a dominant paradigm (Huang et al., 2025; He et al., 2026).

Recent studies have further advanced this direction through multi-modal integration, increasingly expressive network architectures, and large-scale representation learning. MMPD-DTA (Wang et al., 2025a) integrated graph-based and sequence-based representations of proteins and ligands, while T-ALPHA (Kyro et al., 2025) represented pocket surface geometry using oriented point clouds. Architectural innovations, such as CheapNet (Lim et al., 2025) and PLMAM-PLA (Wang et al., 2025b), were designed to emphasize binding-relevant substructures. In parallel, SableBind (Li & Gong, 2025) employed self-supervised structural perturbations to refine molecular embeddings, and ESP-DTI (Wu et al., 2026) integrated a cross-modal contrastive alignment to improve generalization under distribution shifts. Despite these advances, existing methods remain largely biologically agnostic, often treating bioactivity as an isolated ligand–protein property while overlooking the ligand competition mechanism that fundamentally gives rise to bioactivity.

D Additional Dataset Details

Additional details of CompDB are summarized in Table 7. The table reports, for each target class and assay subset, the numbers of assays, samples, proteins, protein structures (derived from either crystal structures or AlphaFold models), binding pockets, radioligand-occupied pockets (together with the number of unique radioligands), and test ligands. Aggregate statistics across all target classes and assays are provided at the end of the table. GPCR and SLC constitute the larger target classes, whereas AR is a small subclass of GPCR and is used for the ablation study.

Table 7: Summary of the proposed CompDB dataset in this study.

Target Class	Assay	#Assays	#Samples	#Proteins	#Pockets	#Radioligand-Occupied Pockets	#Ligands
GPCR	IC_{50}	9,696	13,963	136	185	406	7,999
SLC	IC_{50}	739	1,908	12	22	80	578
AR	IC_{50}	253	610	4	10	24	197
All Targets	All Assays	10,675	16,481	151	214	505	8,674

[a] Samples denote the total number of pocket–radioligand–ligand poses in each subset.
[b] A radioligand-occupied pocket is defined as a binding pocket bound by a radioligand. For each protein, both predefined docking sites and an additional blind-docking setting were considered; in the latter case, the entire protein structure was treated as a single pocket for counting purposes.

E Evaluation Metrics

This section provides formal definitions of all evaluation metrics employed in this study. For clarity and consistency, the notation used in the metric equations is defined at the end of the section and reused throughout.

Root Mean Squared Error (RMSE)

RMSE is calculated as the square root of the average of the squared differences between the actual values and the predicted values. It measures the magnitude of the errors in the same units as the target variable and heavily penalizes large errors (outliers).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

MEAN ABSOLUTE ERROR (MAE)

MAE is the average of the absolute differences between the actual and predicted values. It provides an assessment of average error magnitude, making it more robust to outliers than RMSE since it does not square the errors.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$$

PEARSON'S r (LINEAR CORRELATION COEFFICIENT)

Pearson's r quantifies the strength and direction of the linear relationship between the actual values and the predicted values. It ranges from -1 (perfect negative correlation) to $+1$ (perfect positive correlation).

$$r_{Y,\hat{Y}} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})}{\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}\sqrt{\sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})^2}}$$

SPEARMAN'S ρ (RANK CORRELATION COEFFICIENT)

Spearman's ρ is a non-parametric measure that assesses the monotonic relationship (the consistency of the relationship's direction) between the ranks of the actual values and the ranks of the predicted values.

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

KENDALL'S τ (KENDALL RANK CORRELATION)

Kendall's τ is another non-parametric rank correlation measure based on counting the number of concordant (N_c) and discordant (N_d) pairs to quantify the degree of association between two rankings.

$$\tau = \frac{N_c - N_d}{\frac{1}{2}n(n-1)}$$

PRECISION@K

Precision@k evaluates the enrichment of true actives among the top-ranked predictions and reflects early retrieval performance in virtual screening. It answers the question: among the top k compounds selected according to predicted bioactivity, what fraction are truly active?

Let $\mathcal{I}_k$ denote the set of indices corresponding to the top k predictions ranked by $\hat{Y}_i$ (e.g., highest predicted bioactivity). Let $A_i \in \{0, 1\}$ be a binary indicator denoting whether compound i is experimentally active under a predefined bioactivity threshold. Precision@k is defined as

$$\text{Precision@}k = \frac{1}{k}\sum_{i\in\mathcal{I}_k} A_i.$$

Precision@k ranges from 0 to 1, where higher values indicate better enrichment of active compounds among the top-ranked candidates.

NOTATION KEY

n Total number of data points or observations.

Y_i The i-th actual (observed) value.

$\hat{Y}_i$ The i-th predicted value.

$\bar{Y}$ The mean of the actual values.

$\bar{\hat{Y}}$ The mean of the predicted values.

d_i The difference between the ranks of Y_i and $\hat{Y}_i$ for the i-th data point (used in Spearman's ρ).

N_c The number of concordant pairs.

N_d The number of discordant pairs.

Table 8: Model architecture hyperparameters.

Model	# Layers	# Hidden dim.	# Attn heads[a]	Dropout
HypSeek	3	128	4	0.02
T-ALPHA	4	128	4	0.02
GIGN, EGNN	3	128	N.A.	0.02
CheapNet, DTIGN	3	128	4	0.02
GEMS	2	128	4	0.02

[a] Number of attention heads used inside the model.

Table 9: Node features.

Common node features	Type	Size	Description
Atom type	One-hot	10	Heavy atom type [C, N, O, S, F, P, Cl, Br, I, others]
Degree	One-hot	7	Number of covalent bonds [0, 1, 2, 3, 4, 5, 6]
Implicit valence	One-hot	7	Implicit valence of the atom [0, 1, 2, 3, 4, 5, 6]
Hybridization	One-hot	5	[sp, sp2, sp3, sp3d, sp3d2]
Aromaticity	Binary	1	Whether the atom is part of an aromatic system
Hydrogens	One-hot	5	Number of connected hydrogens [0, 1, 2, 3, 4]
GEMS node features	**Type**	**Size**	**Description**
Atom type	One-hot	9	[B, C, N, O, P, S, Se, metal, halogen]
Ring membership	Binary	1	Whether the atom is in a ring
Hybridization	One-hot	8	[s, sp, sp2, sp2d, sp3, sp3d, sp3d2, unspecified]
Formal charge	Float	1	Formal charge of the atom
Aromaticity	Binary	1	Whether the atom is part of an aromatic system
Atomic mass	Float	1	Atom mass in daltons divided by 100
Hydrogens	Integer	1	Number of bonded hydrogens
Degree	One-hot	10	[0, 1, 2, 3, 4, 5, 6, 7, 8, others]
Chirality	One-hot	10	[unspecified, cw, ccw, others]
Amino acid type (protein-only)	One-hot	20	[ALA, ARG, ASN, ASP, CYS, GLN, GLU, GLY, HIS, ILE, LEU, LYS, MET, PHE, PRO, SER, THR, TRP, TYR, VAL]
Amino acid embedding (protein-only)	Real vector	320	Amino-acid embeddings generated using ESM2-T6-8M, obtained from `huggingface.co/facebook/esm2_t6_8M_UR50D`
Additional node features introduced in DrugComp	**Type**	**Size**	**Description**
Node type	One-hot	3	Node in [protein, radioligand, ligand]
Vina scoring terms	Real vector	5	$[\text{Gauss}_1, \text{Gauss}_2, \text{Repulsion}, \text{Hydrophobic}, \text{HBond}]$ aggregated from neighbor nodes
Atom-level Vina score	Float	1	Weighted sum of Vina scoring terms

F Additional Implementation Details

F.1 Baseline Models

To evaluate the performance of DrugComp, we consider a set of representative protein–ligand binding prediction models spanning transformer-based and graph-based architectures. DrugComp is applied to each of these baseline models to assess whether incorporating competitive binding improves predictive performance. The baseline models are summarized below in chronological order based on publication year:

- **HypSeek** Wang et al. (2026) is a hyperbolic representation learning framework that embeds ligands, protein pockets, and protein sequences into Lorentz-model hyperbolic space through a protein-guided three-tower architecture, enabling affinity-sensitive modeling for virtual screening and affinity ranking.
- **T-ALPHA** Kyro et al. (2025) is a hierarchical transformer model that integrates multimodal features to capture both local and global dependencies within protein–ligand complexes, improving binding affinity prediction accuracy.

Table 10: Edge features.

Common intra-mol edge features	**Type**	**Size**	**Description**
Bond type	One-hot	4	[single, double, triple, aromatic]
Conjugation	Binary	1	Whether the bond is conjugated [0/1]
Ring	Binary	1	Whether the bond is in a ring [0/1]
Stereo	One-hot	4	[StereoNone, StereoAny, StereoZ, StereoE]
DTIGN inter-mol edge features	**Type**	**Size**	**Description**
Distance (d) in Coulomb force	Real vector	64	$\left[\exp\left(-\frac{\|d^{-2}-\phi_k\|^2}{(\phi_{k+1}-\phi_k)^2}\right)\right]_{k=0}^{63}$ where $\phi_k = \left(6-\frac{5k}{63}\right)^{-2}$
Distance (d) in London dispersion	Real vector	64	$\left[\exp\left(-\frac{\|d^{-6}-\phi_k\|^2}{(\phi_{k+1}-\phi_k)^2}\right)\right]_{k=0}^{63}$ where $\phi_k = \left(6-\frac{5k}{63}\right)^{-6}$
GEMS unified edge features	**Type**	**Size**	**Description**
Edge type	One-hot	3	[covalent, self-loop, non-covalent]
Length / Distance	Float	1	Bond length (intra-molecular) or Euclidean distance (inter-molecular)
Bond type	One-hot	4	[single, double, triple, aromatic].
Conjugation	Binary	1	Whether the bond is conjugated [0/1].
Ring membership	Binary	1	Whether the bond is in a ring [0/1].
Stereochemistry	One-hot	6	Stereo type [None, Any, E, Z, CIS, Trans].
Atom-to-residue distances	Real vector	4	Distances between the ligand atom and the four main backbone atoms in the residue (N, C_α, C, virtual C_β).
Additional edge features introduced in DrugComp	**Type**	**Size**	**Description**
Edge type	One-hot	6	[protein-protein, radioligand-radioligand, ligand-ligand, protein-radioligand, protein-ligand, radioligand-ligand]

- **GEMS** Graber et al. (2025) addresses dataset bias in binding affinity prediction by learning more robust protein–ligand representations, enhancing generalization across diverse protein families.
- **CheapNet** Lim et al. (2025) is a cross-attention model operating on hierarchical protein and ligand representations, providing computationally efficient binding affinity prediction while capturing relevant interaction patterns.
- **DTIGN** Yin et al. (2024) integrates interatomic forces from molecular docking with self-attention on intermolecular graphs, explicitly modeling drug–target interactions to improve bioactivity prediction.
- **GIGN** Yang et al. (2023) constructs a geometric interaction graph for protein–ligand complexes, explicitly encoding 3D structural information and interaction features to predict binding affinity.
- **EGNN** Satorras et al. (2021) is an E(n) equivariant graph neural network that produces representations invariant or equivariant to rotations, translations, reflections, and permutations, enabling efficient modeling of geometric data without higher-order tensor representations.

For all baseline models, a consistent multi-layer perceptron (MLP) head is used for final bioactivity prediction, except for GEMS, which retains its original architecture for the output layer. The MLP consists of three fully connected layers, with leaky ReLU activation and 1D batch normalization applied after the first two layers. For HypSeek, the pair-wise attention in the Uni-Core transformer encoder scales substantially with the number of pocket nodes; therefore, for both its baseline and DrugComp-enhanced variants, we restrict the union pocket to local pocket atoms within 6 Å of the ligand or radioligand, and in DrugComp, allow full attention only between ligand and radioligand nodes.

Hyperparameters are tuned individually for each baseline model and summarized in Table 8. Specifically, we search the number of layers from $\{1, 2, 3, 4, 5\}$, hidden dimensions from $\{64, 128, 256\}$, attention heads from $\{1, 2, 4, 8\}$, and dropout rates from $\{0, 0.02, 0.05\}$. Final settings are selected according to the baseline performance on the CompDB-AR test set; when multiple settings yield comparable performance, we choose the one most consistent with the other models. Each DrugComp-enhanced variant uses the same hyperparameter settings as its corresponding baseline model, allowing a controlled evaluation of improvements attributable to the competitive binding framework.

F.2 Graph Features

In our implementation, the graph representation is adapted to the architectural design of each evaluated model. For HypSeek, T-ALPHA, CheapNet, GIGN, and EGNN, we construct atom-level protein pocket-ligand graphs using RDKit v2024.03.5. These models share the same node features and intra-molecular edge features, as summarized in Table 9 (Top) and Table 10 (Top). Building on this common representation, DTIGN further introduces inter-molecular edges derived from Coulomb interactions and London dispersion forces, whose features are also reported in Table 10 (Middle).

In contrast, GEMS follows a different graph representation. It adopts a distinct set of atom features and employs a unified edge representation, where intra-molecular covalent bonds and inter-molecular non-covalent interactions are encoded within the same edge feature space. For ligand representation, GEMS additionally incorporates pretrained ChemBERTa-2 (ChemBERTa-77M-MLM) embeddings obtained from `huggingface.co/DeepChem/ChemBERTa-77M-MLM`, which are concatenated with graph-level ligand features within the network. The corresponding feature definitions are summarized in Table 9 (Bottom) and Table 10 (Bottom).

F.3 Dataset Splits

For all CompDB datasets (CompDB-GPCR, CompDB-SLC, and CompDB-AR), we employ 5-fold cross-validation to evaluate model performance. The test sets across the five folds are non-overlapping and collectively cover all samples in each dataset. Models are trained and evaluated independently on each fold, and final results are reported as the mean and standard deviation of test metrics across folds.

We use the KFold function from the scikit-learn v1.3.0 library with a fixed random seed of 1 for reproducible splitting. The dataset was split at the tuple level based on unique radioligand–ligand–protein (R, L, P) combinations, ensuring that all complexes associated with the same (R, L, P) tuple were assigned to the same subset and preventing information leakage across training and testing folds. Within each training fold, a validation set was further separated from the training data using the train_test_split function (random state = 1) for early stopping.

F.4 Training Details

All models were trained on a Linux workstation equipped with four NVIDIA RTX 4090 GPUs (24 GB memory each). On a single GPU, training DTIGN + DrugComp on the largest CompDB-GPCR dataset ($\sim$10k samples) requires approximately 1 minute per epoch, compared to $\sim$30 seconds for the DTIGN baseline. Training time varies across backbone models. With early stopping typically triggered around 100 epochs, the training time per fold is approximately 2 hours for the largest dataset. All experiments were implemented in Python 3.8.

Model selection was based on validation RMSE. Early stopping with a patience of 30 epochs was applied. All models were optimized using the Adam optimizer with a learning rate of 0.001, except HypSeek using 0.0001 for stable training, and a weight decay of $1 \times 10^{-4.3}$.