TENSOR-DTI: ENHANCING BIOMOLECULAR INTERACTION PREDICTION WITH CONTRASTIVE EMBEDDING LEARNING

Manel Gil-Sorribes[1], **Álvaro Ciudad**[1,2], **Alexis Molina**[1,2*]

[1]Nostrum Biodiscovery, Barcelona, 08029, Spain
[2]ATLAS Labs, Barcelona, Spain
*alexis.molina@nostrumbiodiscovery.com, @theatlaslabs.com

ABSTRACT

Accurate drug-target interaction (DTI) prediction is essential for computational drug discovery, yet existing models often rely on pre-defined molecular descriptors or sequence-based embeddings with limited generalizability. We propose Tensor-DTI, a contrastive learning framework that integrates multimodal embeddings from molecular graphs, protein language models, and binding site predictions to improve interaction modeling. Tensor-DTI employs a Siamese Dual Encoder architecture, enabling it to capture both chemical and structural interaction features while distinguishing interacting from non-interacting pairs. Evaluations on multiple DTI benchmarks, including BIOSNAP, BindingDB, DAVIS, and PLINDER, demonstrate that Tensor-DTI outperforms existing sequence-based and graph-based models. Additionally, we assess its generalization to unseen drugs and proteins and explore its applicability to protein-RNA and peptide-protein interactions. Our findings highlight the benefits of integrating structural information with contrastive objectives to enhance interaction prediction accuracy and model interpretability.

1 INTRODUCTION

The vast chemical space, estimated at up to 10^{60} small molecules (Restrepo, 2022), presents a major challenge for drug discovery, as practical exploration is constrained by synthesizability, stability, and biological relevance. Even after applying physicochemical filters such as Lipinski's Rule of Five (Lipinski et al., 1997), the number of feasible drug-like molecules remains in the range of 10^{12} to 10^{23}, making exhaustive experimental screening impractical. High-throughput screening (HTS) and virtual screening (HTVS) help navigate this space, but HTS is costly and limited to predefined libraries, while HTVS, particularly molecular docking, remains computationally intensive. The purchasable chemical space, bolstered by combinatorial chemistry, has grown exponentially, with ENAMINE REAL (Enamine) containing over 1.4 trillion readily synthesizable compounds and ZINC22 (Tingle et al., 2023) offering access to more than 2 billion purchasable molecules. Despite these advances, screening approaches remain constrained by computational limitations and the reliance on pre-existing receptor-ligand conformations, limiting generalizability.

Machine learning-based approaches have emerged as an alternative, accelerating docking predictions while bypassing the need for exhaustive sampling. Models such as DiffDock (Corso et al., 2023) and TankBind (Lu et al., 2022) leverage transformer-based architectures and generative models to predict binding poses with high efficiency. However, their reliance on large co-crystal datasets presents a bottleneck, as structural data for diverse ligand-protein interactions remain scarce. Molecule-protein interaction datasets, including ChEMBL (Gaulton et al., 2012), PDBBind (Wang et al., 2004), and DUD-E (Mysinger et al., 2012), provide a foundation for training predictive models, but data quality issues and information leakage between training and test sets hinder model robustness (Li et al., 2024). As a result, modern drug-target interaction (DTI) models have focused on leveraging protein language models such as ESM (Rives et al., 2019) and SaProt (Su et al., 2023), as well as graph-based representations like GraphDTA (Nguyen et al., 2021) and HyperAttentionDTI (Zhao et al., 2022), to enhance interaction prediction.

However, current DTI models still face fundamental challenges in capturing the full complexity of biomolecular interactions. Many existing approaches rely on whole-protein embeddings, overlooking the importance of localized binding site information, which plays a critical role in molecular recognition and selectivity. While graph neural networks (GNNs) and transformer-based architectures have improved interaction modeling (Tsubaki et al., 2019; Chen et al., 2020), they often struggle to generalize to unseen drugs and targets due to their dependence on fixed molecular representations (Yu et al., 2012). Contrastive learning frameworks such as ConPLex (Singh et al., 2023) and PocketDTA (Zhao et al., 2024) have attempted to refine feature spaces by embedding proteins and drugs in a shared representation, but most fail to explicitly incorporate multimodal structural information, which is essential for capturing the nuances of binding affinities and selectivity. Additionally, generalization remains a major concern, as existing models often perform poorly on interactions beyond their training distribution, limiting their real-world applicability.

Building on these advances, we introduce Tensor-DTI, a deep learning framework that integrates multimodal embeddings and a Siamese Dual Encoder with contrastive learning to enhance DTI prediction. Unlike traditional sequence-based models, Tensor-DTI incorporates both global and localized structural features, leveraging pocket embeddings alongside protein and ligand representations to refine binding-site specificity. The model combines structural, chemical, and contextual information, enabling it to generalize across diverse biomolecular interactions, including peptide-protein and RNA-protein binding. By explicitly modeling binding pockets using ESM-GearNet (Zhang et al., 2023a), a hybrid approach that integrates protein language models with structural message-passing networks (Zhang et al., 2023b), Tensor-DTI provides a more interpretable and biologically relevant representation of molecular interactions. This architecture allows Tensor-DTI to outperform existing models on standard benchmarks while offering insights into interaction specificity, making it a scalable and generalizable tool for drug discovery.

2 TENSOR-DTI

Tensor-DTI is a deep learning framework for DTI prediction that integrates multimodal molecular representations with contrastive learning. The model employs a Siamese Dual Encoder architecture, where separate encoder branches process drug and protein representations, projecting them into a shared latent space. A contrastive loss function encourages the embeddings of interacting pairs to cluster while pushing non-interacting pairs apart, enabling generalization to unseen drug-target pairs. For binary interaction classification, a binary cross-entropy loss is applied, ensuring the model learns a probabilistic interaction score.

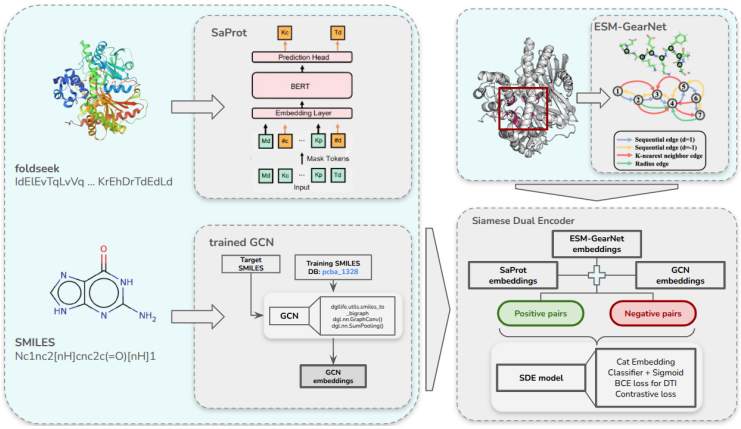


Figure 1: Tensor-DTI architecture with pocket embeddings. The model extends the base architecture by incorporating binding pocket representations, enabling site-specific interaction modeling. The protein shown is PDBID: 5ISX. The SaProt image is adapted from (Su et al., 2023), and the GearNet image is adapted from (Zhang et al., 2023b).

Small molecules are represented as molecular graphs, and Tensor-DTI extracts drug embeddings using a Graph Convolutional Network (GCN) trained on PCBA_1328, a dataset of 1.6M molecules

with 1,328 binary activity labels from PubChem (Kim et al., 2023). The GCN iteratively aggregates local structural features, encoding bioactivity-relevant molecular patterns into a single graph-level embedding via sum pooling. Proteins are primarily represented using transformer-based embeddings from SaProt, a language model that integrates sequence and structural information. In a specific case, we also used ESM-2, which encodes high-resolution sequence features from large-scale protein corpora. These embeddings provide rich contextualized representations of proteins, enabling the model to learn functionally relevant interaction patterns.

To refine binding-site specificity, Tensor-DTI incorporates pocket embeddings, allowing it to distinguish between global protein interactions and site-specific binding events. These embeddings are derived using ESM-GearNet (Zhang et al., 2023a), which refines protein language model embeddings with structural information via GearNet (Zhang et al., 2023b), a graph-based message-passing network that captures residue-residue interactions. To enhance spatial awareness, a multi-layer perceptron is trained on the SC-PDB dataset (Kellenberger et al., 2006) for node classification, enabling the model to differentiate functionally relevant binding sites from non-binding regions. Pocket embeddings are combined with full-protein representations, ensuring that the model effectively captures ligand-pocket interactions while maintaining contextual protein information. A full visualization of the model architecture, including the integration of pocket embeddings, is provided in Appendixes A and B.

Beyond small-molecule interactions, Tensor-DTI extends to RNA-protein and peptide-protein interactions, broadening its applicability to biomolecular modeling. Peptide representations are extracted from PeptideBERT (Guntuboina et al., 2023), a transformer-based model trained on peptide sequences, while RNA embeddings are generated using ChaRNABERT (Morales-Pastor et al., 2024), which employs gradient-based subword tokenization to dynamically segment RNA sequences, capturing both nucleotide-level interactions and higher-order structural dependencies. These additional representations allow the model to generalize beyond conventional drug-protein interactions and accommodate alternative therapeutic modalities.

For drug-target affinity (DTA) prediction, Tensor-DTI is adapted to a regression framework by replacing contrastive loss with mean squared error loss. This modification allows the model to predict continuous affinity values rather than binary interactions. Drug and protein embeddings remain consistent with those used in classification tasks, ensuring a unified approach across predictive settings. The model is optimized using the Adam optimizer, with early stopping applied to prevent overfitting based on validation performance. By leveraging contrastive learning for classification and adapting seamlessly to affinity prediction, Tensor-DTI provides a flexible and scalable approach for modeling molecular interactions across diverse biological contexts. Details on the hyperparameters used for different settings can be found in Appendix C.

Additionally, we develop a confidence model, which, although not used for the presented benchmarks, we consider essential for applying Tensor-DTI in prospective scenarios where ground truth data is unavailable. See Appendix D for further details.

3 Results and Discussion

We evaluate Tensor-DTI across multiple DTI benchmarks, comparing it to competitive methods in both classification and affinity prediction tasks. Additionally, we assess its prospective applicability in real-world inference scenarios using recent leak-proof datasets.

Ablation studies (see Appendix E) identified pretrained GCN for drugs and SaProt for proteins as the best combination for DTI, while the optimal embeddings for DTA varied by dataset. A full description of all datasets, including preprocessing, splitting strategies, and dataset-specific details, is in Appendix F, while dataset volumes are in Appendix G.

3.1 Benchmarking Tensor-DTI

To evaluate the predictive performance of Tensor-DTI in DTI modeling, we conducted benchmarking experiments on multiple standard datasets, including BIOSNAP, BindingDB, and DAVIS, alongside two additional BIOSNAP splits assessing generalization to unseen drugs and unseen targets. The results, summarized in Table 1, provide a comparative assessment against established deep learning baselines and classical machine learning methods.

Table 1: Model Performance on standard DTI datasets. Each model for each dataset has been run 5 times. Performance is reported as the Area Under Precision Recall (AUPR) of the prediction. Metrics for models with † are taken from ref. (Huang et al., 2020). Ridge regression is not applicable to the Unseen Drugs dataset split because a distinct model is trained for each drug in the training set.

Model	**BIOSNAP**	**BindingDB**	**DAVIS**	**Unseen Drugs**	**Unseen Targets**
Tensor-DTI	0.903 ± 0.003	0.699 ± 0.002	0.547 ± 0.006	0.888 ± 0.002	0.839 ± 0.003
ConPLex	0.897 ± 0.001	0.628 ± 0.012	0.458 ± 0.016	0.874 ± 0.002	0.842 ± 0.006
EnzPred-CPI	0.866 ± 0.003	0.602 ± 0.006	0.277 ± 0.009	0.844 ± 0.005	0.795 ± 0.004
MolTrans	0.885 ± 0.005	0.598 ± 0.013	0.335 ± 0.017	0.863 ± 0.005	0.668 ± 0.045
GNN-CPI†	0.890 ± 0.004	0.578 ± 0.015	0.269 ± 0.020	–	–
DeepConv-DTI†	0.889 ± 0.005	0.611 ± 0.015	0.299 ± 0.039	0.847 ± 0.009	0.766 ± 0.022
Ridge	0.641 ± 0.000	0.516 ± 0.000	0.320 ± 0.000	N/A	0.617 ± 0.000

Tensor-DTI achieved the highest predictive performance across all datasets, with mean AUPR scores of 0.903 ± 0.003 on BIOSNAP, 0.699 ± 0.002 on BindingDB, and 0.547 ± 0.006 on DAVIS. Notably, BIOSNAP exhibits a relatively well-characterized interaction landscape, primarily comprising known, high-confidence drug-target interactions. The model's superior performance on BIOSNAP suggests its ability to effectively capture high-level interaction patterns, likely facilitated by contrastive embedding learning, which optimizes the separation of interacting and non-interacting pairs.

The more challenging BindingDB dataset, which encompasses a broader range of experimentally validated interactions across diverse molecular classes, results in lower overall predictive performance for all models. Tensor-DTI maintains a robust performance margin over alternative deep learning models such as ConPLex ($+7.1\%$), MolTrans (Huang et al., 2020) ($+10.1\%$), and EnzPred-CPI (Goldman et al., 2022) ($+9.7\%$). The relatively lower performance on DAVIS, where binding interactions are limited to kinase inhibitors, highlights the inherent challenge of predicting selective interactions within structurally conserved protein families.

Among competing methods, ConPLex performs well on BIOSNAP (0.897 ± 0.001) but exhibits a significant drop on BindingDB (0.628 ± 0.012) and DAVIS (0.458 ± 0.016), suggesting a sensitivity to data heterogeneity and potential limitations in generalization beyond the training distribution. EnzPred-CPI and MolTrans show comparatively lower performance, particularly on DAVIS (0.277 and 0.335, respectively), where kinase inhibitors exhibit complex binding profiles that are difficult to capture with purely sequence-based representations. Ridge regression, as expected, exhibits the lowest performance across all datasets, reinforcing the necessity of deep-learning-based representations for capturing the non-linear and high-dimensional features governing biomolecular interactions.

Beyond in-distribution benchmarking, we assessed Tensor-DTI's capacity to generalize to novel drug-like molecules and previously unobserved protein targets. The unseen drug split evaluates the model's ability to infer interactions for chemical entities that do not appear in the training set, whereas the unseen target split assesses generalization to proteins with no direct training exposure.

Tensor-DTI exhibits superior performance in the unseen drug scenario, with an AUPR score of 0.888 ± 0.002, and achieves 0.839 ± 0.003 in the unseen target scenario, showing comparable performance to ConPLex (0.842 ± 0.006) as the difference lies within the margin of error. These results indicate that Tensor-DTI captures meaningful chemical and protein features, enabling it to extend beyond memorized interactions. ConPLex follows with 0.874 ± 0.002 for unseen drugs, further highlighting its competitive performance in generalization tasks.

MolTrans and DeepConv-DTI (Lee et al., 2019) show greater variability, particularly in the unseen target setting (0.668 ± 0.045 and 0.766 ± 0.022, respectively), suggesting higher sensitivity to dataset distribution shifts. The relatively lower performance of MolTrans across both unseen drug and target scenarios underscores the challenge of extrapolating to novel chemical or protein scaffolds.

3.2 Effectiveness of the Contrastive Learning Approach and Evaluation on DUD-E Dataset

To further assess the effectiveness of the contrastive learning approach employed in Tensor-DTI, we evaluated the model on the DUD-E dataset (Mysinger et al., 2012), focusing on the kinase family.

The performance of Tensor-DTI on this task is illustrated through a t-SNE visualization of the learned embeddings, with an example for one of the test proteins shown in Figure 2. Prior to contrastive training, the embeddings of proteins and drugs lack clear separation between actives and decoys. After training, the model successfully clusters active drugs closer to their corresponding protein targets in the latent space, demonstrating improved discrimination between true binders and decoys. This structured embedding space suggests that the model effectively captures interaction-relevant molecular features. Tensor-DTI achieved an average AUPR of 0.686 ± 0.006, for all the test set, across five independent executions, confirming its strong capability in distinguishing actives from decoys.

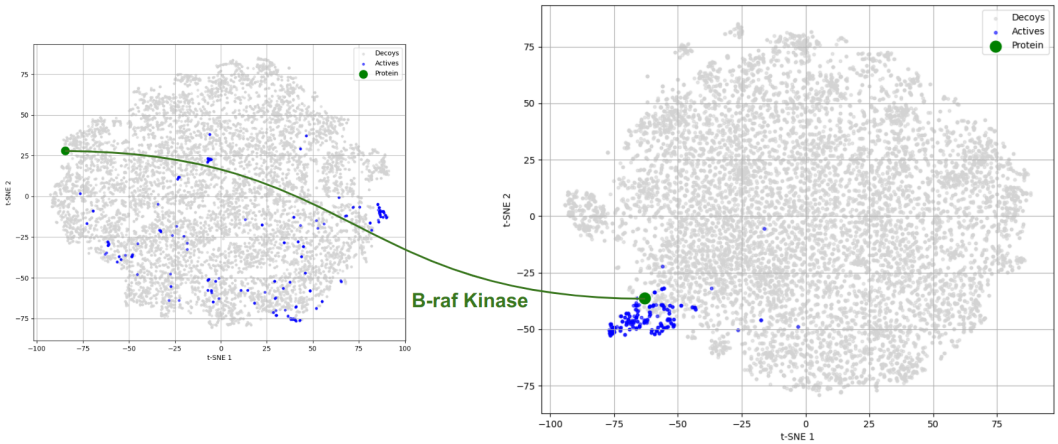


Figure 2: t-SNE visualization of protein and drug embeddings before (left plot) and after (right plot) applying Tensor-DTI with contrastive learning. The visualization corresponds to the B-raf Kinase protein, one of the targets from the test split.

These results reinforce the generalizability of Tensor-DTI, showing that contrastive learning not only improves predictive performance but also enhances interpretability. The structured latent space could be further leveraged for generative modeling, enabling the exploration of new active compounds based on proximity in the learned representation space.

3.3 Tensor-DTI Performs Well on Affinity Prediction Datasets

We tested Tensor-DTI on the Therapeutics Data Commons (TDC) DTI Domain Generalization (Huang et al., 2021) (TDC-DG) benchmark, a challenging dataset for binding affinity prediction. The benchmark includes IC50 values from interactions patented between 2013 and 2018 as training data, while test interactions are from patents filed in 2019 and 2021. This setup demands strong out-of-domain generalization, simulating real-world applications where models predict unseen interactions based on historical data. Following the data split strategy outlined in (Singh et al., 2023), we trained and evaluated Tensor-DTI, which achieved a Pearson Correlation Coefficient (PCC) of 0.580 ± 0.004, demonstrating that our model is competitive with several state-of-the-art methods (Table 2).

To ensure robust performance, we experimented with multiple molecular and protein representations. Our findings showed that using Morgan Fingerprints (MFPS) for small molecules and ESM-2 embeddings for proteins yielded the best results (see Appendix E).

3.4 DTI and DTA Assessments on Low-Leakage Datasets

In order to evaluate drug–target interaction and affinity prediction models under minimized data leakage, we assessed performance across two curated datasets: PLINDER (Durairaj et al., 2024) and

Table 2: Comparison of Tensor-DTI performance on the TDC-DG benchmark.

Model	**PCC**
Tensor-DTI	0.580 ± 0.004
ConPLex	0.538 ± 0.008
MMD	0.433 ± 0.010
CORAL	0.432 ± 0.010
ERM	0.427 ± 0.012
MTL	0.425 ± 0.010

LP-PDBBind (Li et al., 2024). These datasets were designed to reduce structural redundancy and prevent information leakage between training and test sets, making them valuable for assessing the generalization capacity of modern predictive models. Performance comparisons across these datasets are provided in Appendix H.

For PLINDER, we conducted two classification-based evaluations with different negative sampling strategies. In the first split, using only drug and protein embeddings, negative pairs were randomly selected from the same pool of drugs and proteins within each respective split, ensuring that only molecules present in that specific split were used for constructing non-interacting pairs. This approach resulted in a classification AUPR of 0.779 ± 0.001. In contrast, a more challenging split, which incorporated pocket embeddings and enforced high pocket dissimilarity for negative examples, resulted in a reduced AUPR of 0.698 ± 0.005. This discrepancy highlights a key observation: many existing DTI models implicitly rely on pocket similarity rather than learning fundamental binding principles. The more diverse pocket-based negatives force the model to differentiate interactions beyond simple sequence or structure similarity, revealing potential overfitting issues in conventional approaches.

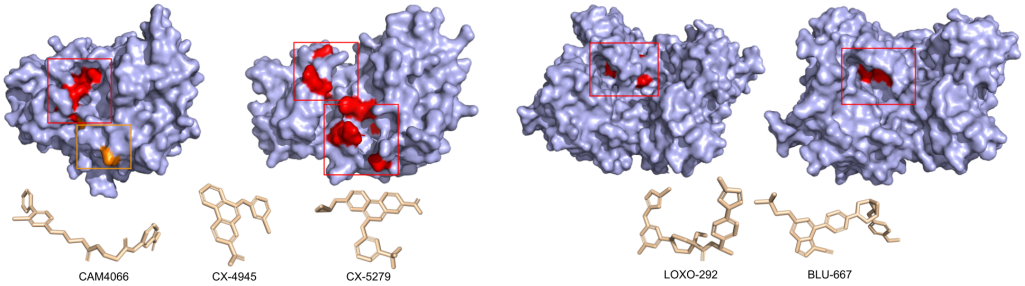
For LP-PDBBind, which focuses on binding affinity prediction, we tested the model respecting the splits proposed by the authors in (Li et al., 2024). Our model, which predicts the K_d, resulted in a PCC of 0.565 ± 0.004 with a Root Mean Square Error (RMSE) of 1.620 ± 0.024. The model achieved lower PCC (0.528 ± 0.013) and higher RMSE (2.122 ± 0.032) for ΔG than for K_d, indicating greater noise and complexity in free energy predictions. To further investigate ligand-specific effects, we used the PDBBind-Opt (Wang et al., 2024) dataset separately to peptides and small molecules. For peptides, the PCC reached 0.679 ± 0.014, with a corresponding RMSE of 1.175 ± 0.020. Meanwhile, for molecule-protein interactions, Tensor-DTI achieved a PCC of 0.750 ± 0.005 with a RMSE of 1.335 ± 0.011 on a random PDBBind-Opt split. However, when evaluated on Li et al. (2024), which ensures no structural or sequence leakage between training and test sets, performance for molecule-protein interactions decreased to a PCC of 0.493 ± 0.005 with a RMSE of 1.545 ± 0.006. This decline reflects the increased difficulty of generalizing to entirely novel protein-ligand interactions, emphasizing the importance of evaluating models under challenging, low-leakage conditions.

3.5 Tensor-DTI Allows Pocket Specificity with the Addition of Pocket Embeddings

Traditional drug discovery often targets entire proteins, but modulating specific binding sites, including allosteric and cryptic pockets, offers greater therapeutic potential. Tensor-DTI enhances site-specific predictions by integrating pocket embeddings derived from ESM-GearNet trained on binding site data, which refines protein binding site representations using ESM-2 embeddings and GearNet-based structural message passing. These embeddings are concatenated with full-protein representations, improving the model's ability to differentiate functionally relevant binding interactions from nonspecific contacts.

3.5.1 Assessment of ATP and AMP Binding Alongside Cryptic Site Inhibitor Binding Predictions in CDK2α and RET Kinases

CDK2 is a key regulator of the G1-to-S phase transition and is frequently hyperactivated in cancers such as breast, ovarian, and certain leukemias (Knudsen et al., 2022). Traditional ATP-competitive in-



(a) *Left.* CDK2 with ATP binding site (red) and closed cryptic site (orange). *Right.* Open cryptic cavity merging with ATP binding site (red). Structural configuration of the three binding compounds to the cryptic site.

(b) *Left.* RET with cryptic binding site in apo state (red). *Right.* Open cryptic cavity (red). Structural configuration of the two binding compounds to the cryptic site.

Figure 3: Structural arrangements from CDK2 and RET kinases in holo and apo states for their correspondent cryptic pockets.

hibitors struggle with selectivity and resistance mechanisms, making alternative binding site targeting an attractive approach. Cryptic binding sites (CBSs) offer a promising avenue for kinase inhibitor development, particularly in overcoming resistance to conventional ATP-competitive inhibitors (Figure 3a). To evaluate Tensor-DTI's ability to distinguish cryptic sites from canonical ATP-binding pockets, we assessed ATP and two well-characterized CBS inhibitors, CX-4945 and CX-5279, across multiple conformational states of CDK2. While the model effectively differentiated between closed ATP sites and open cryptic sites, inconsistencies in ATP binding predictions across cryptic conformations suggested a bias in generalizing ATP's flexibility in alternative binding modes.

The model correctly rejected ATP binding to the closed ATP site in 3FWQ, reinforcing its ability to recognize steric constraints. However, in cryptic conformations where CAM4066 had been removed (5CLP and 5CU3), ATP predictions were inconsistent: while ATP was correctly classified as a binder in 5CLP, it was incorrectly rejected in 5CU3, indicating a limitation in recognizing subtle structural changes affecting pocket accessibility. This discrepancy may stem from a training bias favoring ATP-competitive interactions. In contrast, CX-4945 and CX-5279 were consistently classified as cryptic site binders across CBS structures, highlighting their adaptability and the model's ability to capture non-ATP-competitive interactions. The results align with experimental findings that cryptic site inhibitors generalize better across structural variations than ATP, underscoring their potential in kinase-targeted therapies, particularly for overcoming resistance to ATP-competitive drugs.

Rearranged during transfection (RET) kinase is a critical target in thyroid and lung adenocarcinoma (Liu et al., 2020), but despite the approval of multi-tyrosine kinase inhibitors (MKIs) such as LOXO-292 (selpercatinib) and BLU-667 (pralsetinib), their long-term efficacy is often hindered by secondary mutations, off-target toxicity, and acquired resistance. To address these challenges, researchers have identified a cryptic binding site near the active site as a promising target for next-generation inhibitors (Figure 3b). We evaluated Tensor-DTI's ability to differentiate between the active site and CBS by predicting the binding of AMP, LOXO-292, and BLU-667 across multiple RET conformations. While the model correctly identified cryptic site inhibitors, inconsistencies in ATP-competitive binding predictions suggested areas for improvement in handling binding site transitions.

The model struggled to correctly classify AMP, a known binder to the RET active site (2IVS), suggesting a potential underestimation of ATP-competitive interactions when the CBS is closed. However, it accurately predicted LOXO-292 and BLU-667 binding to the CBS in an open cryptic conformation (7JU5), demonstrating reliable recognition of cryptic site accessibility. Interestingly, in 2IVS, where the active site is open but the CBS remains closed, the model incorrectly predicted CBS binding for these inhibitors, indicating an overgeneralization of cryptic site accessibility. These discrepancies suggest a model bias favoring cryptic site interactions over ATP-competitive binding, highlighting the need for refinements in understanding structural transitions between active and cryptic conformations (Yang et al., 2023).

3.6 Broadening the Scope of Biomolecular Interaction Predictions

Beyond small molecules, Tensor-DTI models peptide-protein, protein-RNA, and drug-RNA interactions, expanding its applicability to biologics and RNA therapeutics. For peptide-protein interactions, Tensor-DTI captures the physicochemical and sequence-dependent features governing peptide binding, achieving an AUPR of 0.953 ± 0.001 on the Propedia (Martins et al., 2023) dataset (Table 14 in Appendix I).

Similarly, for protein-RNA interactions, which are central to post-transcriptional regulation, the model achieves an AUPR of 0.916 ± 0.008 on CoPRA (Han et al., 2024) (Table 15 in Appendix I). When evaluated on PRA310, which provides affinity measurements for protein-RNA pairs, Tensor-DTI achieves a PCC of 0.631 ± 0.111 for K_d (binding constant) and 0.621 ± 0.052 for ΔG (free energy), with corresponding RMSE values of 1.443 ± 0.232 and 1.910 ± 0.212 (Table 16 in Appendix I). While a one-hot encoding baseline performed similarly in RMSE, Tensor-DTI exhibited stronger correlation with true affinities, indicating better predictive accuracy.

For drug-RNA interactions, Tensor-DTI was trained on drug-RNA pairs from PDBBind, achieving a PCC of 0.792 ± 0.015 and an RMSE of 1.684 ± 0.038, outperforming the one-hot encoding baseline, which obtained a PCC of 0.633 ± 0.018 and an RMSE of 1.738 ± 0.036 (Table 17 in Appendix I). Although RMSE values remained comparable, Tensor-DTI's higher PCC suggests superior learning of structure-function relationships, capturing meaningful interaction patterns that conventional encoding methods fail to generalize.

These results demonstrate Tensor-DTI's ability to generalize beyond small-molecule interactions, making it a versatile tool for modeling peptide and RNA interactions in therapeutic applications.

4 Conclusion

Accurate drug-target interaction prediction remains a challenge in computational drug discovery, requiring models that effectively capture the biochemical and structural determinants of molecular recognition. Tensor-DTI enhances DTI prediction by integrating multimodal embeddings from molecular graphs, protein language models, and binding site predictions within a contrastive learning framework. By leveraging a Siamese Dual Encoder architecture, Tensor-DTI not only improves predictive accuracy across diverse DTI benchmarks, including BIOSNAP, BindingDB, DAVIS, and PLINDER, but also demonstrates strong generalization to unseen drugs and proteins.

A key strength of Tensor-DTI is its explicit incorporation of pocket embeddings, refining binding-site specificity and offering a structured, interpretable approach that moves beyond traditional sequence-based DTI models. By capturing both global and localized molecular features, the model enhances interaction modeling while maintaining flexibility for different molecular modalities. The inclusion of contrastive learning objectives ensures the formation of robust, generalizable representations, reducing the risk of overfitting and improving performance in low-data scenarios.

Beyond small-molecule interactions, Tensor-DTI exhibits versatility in biomolecular interaction modeling, extending its applicability to peptide-protein and RNA-associated interactions. This broader scope makes Tensor-DTI particularly valuable for structure-based drug design, RNA therapeutics, and the development of therapeutic agents beyond small molecules. Furthermore, its ability to operate without reliance on explicit structural interaction data enables its use in large-scale virtual screening pipelines, where conventional docking methods remain computationally prohibitive.

Overall, Tensor-DTI represents a scalable and generalizable framework for interaction modeling, balancing accuracy, interpretability, and computational efficiency. Future work will focus on refining its ability to model multi-specific interactions and integrating active learning strategies to further improve predictive robustness and real-world applicability.

MEANINGFULNESS STATEMENT

A meaningful representation of life in DTI modeling should reflect the structural, biochemical, and contextual determinants that drive molecular recognition. Tensor-DTI advances this direction by integrating contrastive multimodal embeddings from molecular graphs, protein language models, and binding site representations, allowing for precise and generalizable interaction modeling. By encoding both global and pocket-level features, Tensor-DTI moves beyond sequence-level heuristics, offering a scalable framework that enhances binding site specificity and interaction prediction. This approach not only improves DTI accuracy but also deepens our understanding of molecular interactions, accelerating data-driven drug discovery and rational molecular design.

REFERENCES

Lifan Chen, Xiaoqin Tan, Dingyan Wang, Feisheng Zhong, Xiaohong Liu, Tianbiao Yang, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, and Mingyue Zheng. TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics*, 36(16):4406–4414, 2020.

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi S. Jaakkola. DiffDock: Diffusion steps, twists, and turns for molecular docking. In *The Eleventh International Conference on Learning Representations*, 2023.

Janani Durairaj, Yusuf Adeshina, Zhonglin Cao, Xuejin Zhang, Vladas Oleinikovas, Thomas Duignan, Zachary McClure, Xavier Robin, Gabriel Studer, Daniel Kovtun, Emanuele Rossi, Guoqing Zhou, Srimukh Veccham, Clemens Isert, Yuxing Peng, Prabindh Sundareson, Mehmet Akdel, Gabriele Corso, Hannes Stärk, Gerardo Tauriello, Zachary Carpenter, Michael Bronstein, Emine Kucukbenli, Torsten Schwede, and Luca Naef. PLINDER: The protein-ligand interactions dataset and evaluation resource. *bioRxiv 2024.07.17.603955*, 2024.

Enamine. Enamine REAL Space. `https://enamine.net/compound-collections/real-compounds/real-space-navigator`. Accessed on November 01, 2024.

Ben Finkelshtein, Xingyue Huang, Michael Bronstein, and İsmail İlkan Ceylan. Cooperative graph neural networks. *arXiv preprint arXiv:2310.01267*, 2023.

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.

Samuel Goldman, Ria Das, Kevin K. Yang, and Connor W. Coley. Machine learning modeling of family wide enzyme-substrate specificity screens. *PLOS Computational Biology*, 18(2):1–20, 02 2022.

Chakradhar Guntuboina, Adrita Das, Parisa Mollaei, Seongwon Kim, and Amir Barati Farimani. PeptideBERT: A language model based on transformers for peptide property prediction. *The Journal of Physical Chemistry Letters*, 14(46):10427–10434, 2023.

Rong Han, Xiaohong Liu, Tong Pan, Jing Xu, Xiaoyu Wang, Wuyang Lan, Zhenyu Li, Zixuan Wang, Jiangning Song, Guangyu Wang, and Ting Chen. CoPRA: Bridging cross-domain pretrained sequence models with complex structures for protein-RNA binding affinity prediction. *arXiv preprint arXiv:2409.03773*, 2024.

Kexin Huang, Cao Xiao, Lucas M Glass, and Jimeng Sun. MolTrans: Molecular interaction transformer for drug–target interaction prediction. *Bioinformatics*, 37(6):830–836, 10 2020.

Kexin Huang, Tianfan Fu, Wenhao Gao, Yue Zhao, Yusuf H Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, Jimeng Sun, and Marinka Zitnik. Therapeutics data commons: Machine learning datasets and tasks for drug discovery and development. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*, 2021.

Esther Kellenberger, Pascal Muller, Claire Schalon, Guillaume Bret, Nicolas Foata, and Didier Rognan. sc-pdb: an annotated database of druggable binding sites from the protein data bank. *Journal of Chemical Information and Modeling*, 46(2):717–727, Mar 2006.

Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, et al. PubChem 2023 update. *Nucleic Acids Research*, 51 (D1):D1373–D1380, 2023.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Erik S Knudsen, Vishnu Kumarasamy, Ram Nambiar, Joel D Pearson, Paris Vail, Hanna Rosenheck, Jianxin Wang, Kevin Eng, Rod Bremner, Daniel Schramek, et al. Cdk/cyclin dependencies define extreme cancer cell-cycle heterogeneity and collateral vulnerabilities. *Cell reports*, 38(9), 2022.

Ingoo Lee, Jongsoo Keum, and Hojung Nam. DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS computational biology*, 15(6): e1007129, 2019.

Jie Li, Xingyi Guan, Oufan Zhang, Kunyang Sun, Yingze Wang, Dorian Bagni, and Teresa Head-Gordon. Leak proof PDBBind: A reorganized dataset of protein-ligand complexes for more generalizable binding affinity prediction. *arXiv preprint arXiv:2308.09639*, 2024.

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

Christopher A. Lipinski, Franco Lombardo, Beryl W. Dominy, and Paul J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23(1):3–25, 1997.

Xuan Liu, Xueqing Hu, Tao Shen, Qi Li, Blaine HM Mooers, and Jie Wu. Ret kinase alterations in targeted cancer therapy. *Cancer Drug Resistance*, 3(3):472, 2020.

Wei Lu, Qifeng Wu, Jixian Zhang, Jiahua Rao, Chengtao Li, and Shuangjia Zheng. TANKBind: Trigonometry-aware neural networks for drug-protein binding structure prediction. *Advances in Neural Information Processing Systems*, 35:7236–7249, 2022.

Pedro Martins, Diego Mariano, Frederico Chaves Carvalho, Luana Luiza Bastos, Lucas Moraes, Vivian Paixão, and Raquel Cardoso de Melo-Minardi. Propedia v2. 3: A novel representation approach for the peptide-protein interaction database using graph-based structural signatures. *Frontiers in Bioinformatics*, 3:1103103, 2023.

Adrián Morales-Pastor, Raquel Vázquez-Reza, Miłosz Wieczór, Clàudia Valverde, Manel Gil-Sorribes, Bertran Miquel-Oliver, Álvaro Ciudad, and Alexis Molina. Character-level tokenizations as powerful inductive biases for RNA foundational models. *arXiv preprint arXiv:2411.11808*, 2024.

Michael M Mysinger, Michael Carchia, John J Irwin, and Brian K Shoichet. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *Journal of Medicinal Chemistry*, 55(14):6582–6594, 2012.

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.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pp. 8748–8763. PMLR, 2021.

Guillermo Restrepo. Chemical space: limits, evolution and modelling of an object bigger than our universal library. *Digital Discovery*, 1(5):568–585, 2022.

Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C. Lawrence Zitnick, Jerry Ma, and Rob Fergus. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *PNAS*, 2019.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010.

Rohit Singh, Samuel Sledzieski, Bryan Bryson, Lenore Cowen, and Bonnie Berger. Contrastive learning in protein language space predicts interactions between drugs and protein targets. *Proceedings of the National Academy of Sciences*, 120(24):e2220778120, 2023.

Jin Su, Chenchen Han, Yuyang Zhou, Junjie Shan, Xibin Zhou, and Fajie Yuan. SaProt: Protein language modeling with structure-aware vocabulary. *bioRxiv 2023.10.01.560349*, 2023.

Benjamin I. Tingle, Khanh G. Tang, Mar Castanon, John J. Gutierrez, Munkhzul Khurelbaatar, Chinzorig Dandarchuluun, Yurii S. Moroz, and John J. Irwin. ZINC-22 A free multi-billion-scale database of tangible compounds for ligand discovery. *Journal of Chemical Information and Modeling*, 63(4):1166–1176, Feb 2023.

Masashi Tsubaki, Kentaro Tomii, and Jun Sese. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*, 35(2):309–318, 2019.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

Renxiao Wang, Xueliang Fang, Yipin Lu, and Shaomeng Wang. The PDBbind database: Collection of binding affinities for protein- ligand complexes with known three-dimensional structures. *Journal of Medicinal Chemistry*, 47(12):2977–2980, 2004.

Yingze Wang, Kunyang Sun, Jie Li, Xingyi Guan, Oufan Zhang, Dorian Bagni, and Teresa Head-Gordon. PDBBind optimization to create a high-quality protein-ligand binding dataset for binding affinity prediction. *arXiv preprint arXiv:2411.01223*, 2024.

Wei-Cheng Yang, Dao-Hong Gong, Hong Wu, Yang-Yang Gao, and Ge-Fei Hao. Grasping cryptic binding sites to neutralize drug resistance in the field of anticancer. *Drug Discovery Today*, 28(9): 103705, 2023.

Hua Yu, Jianxin Chen, Xue Xu, Yan Li, Huihui Zhao, Yupeng Fang, Xiuxiu Li, Wei Zhou, Wei Wang, and Yonghua Wang. A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data. *PloS one*, 7(5):e37608, 2012.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, pp. 12310–12320. PMLR, 2021.

Zuobai Zhang, Chuanrui Wang, Minghao Xu, Vijil Chenthamarakshan, Aurelie Lozano, Payel Das, and Jian Tang. A systematic study of joint representation learning on protein sequences and structures. *arXiv preprint arXiv:2303.06275*, 2023a.

Zuobai Zhang, Minghao Xu, Arian Rokkum Jamasb, Vijil Chenthamarakshan, Aurelie Lozano, Payel Das, and Jian Tang. Protein representation learning by geometric structure pretraining. In *The Eleventh International Conference on Learning Representations*, 2023b.

Long Zhao, Hongmei Wang, and Shaoping Shi. PocketDTA: an advanced multimodal architecture for enhanced prediction of drug- target affinity from 3D structural data of target binding pockets. *Bioinformatics*, 40(10):btae594, 2024.

Qichang Zhao, Haochen Zhao, Kai Zheng, and Jianxin Wang. HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics*, 38(3):655–662, 2022.

A DTI Model Visualization

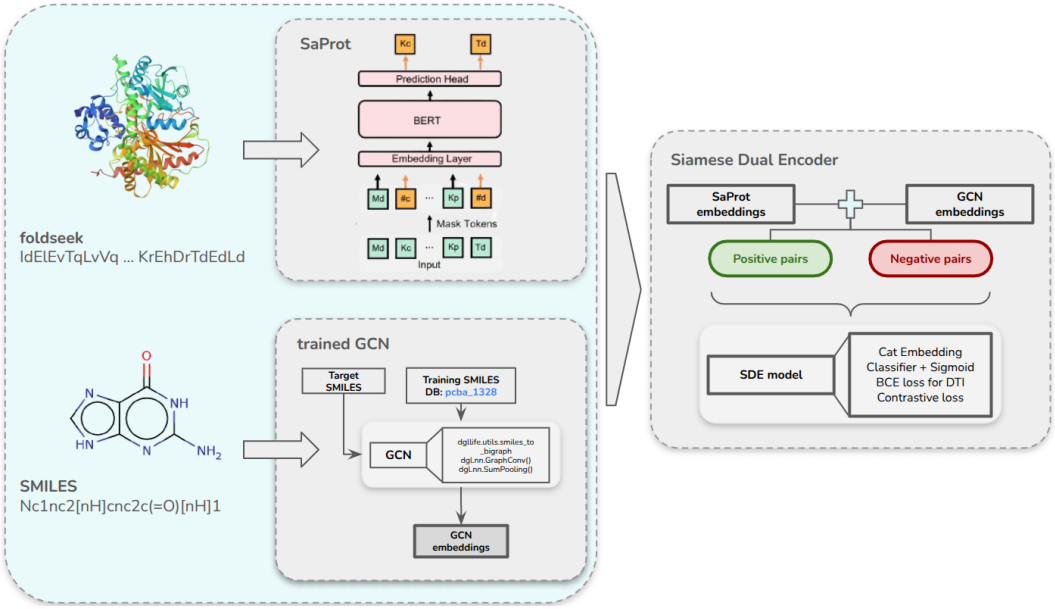


Figure 4: Tensor-DTI architecture. A Siamese Dual Encoder processes multimodal embeddings from drugs and proteins, using contrastive learning to refine the interaction space. The protein shown is 5ISX. The SaProt image is adapted from (Su et al., 2023).

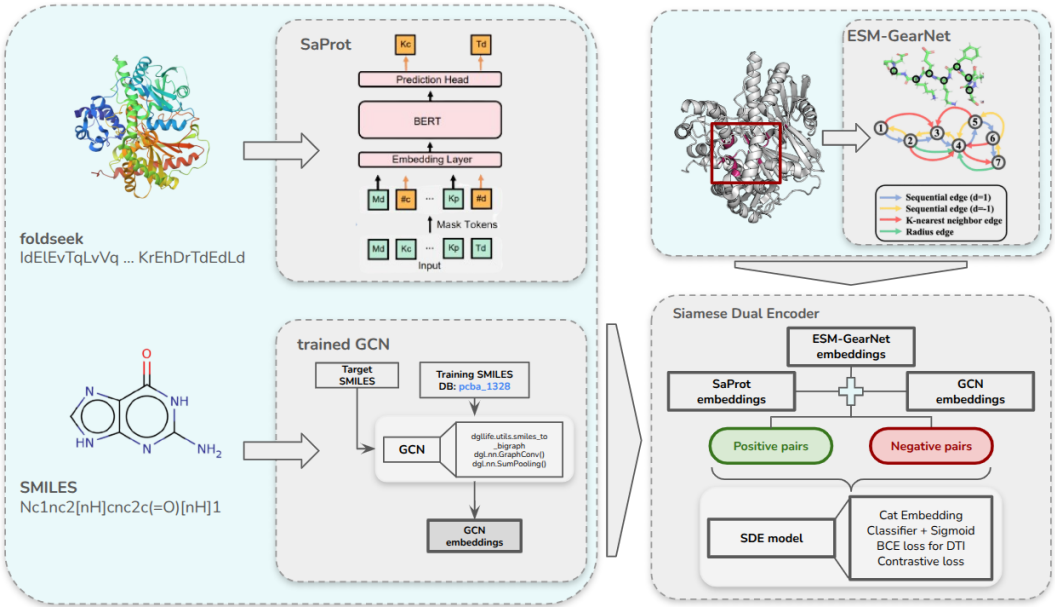


Figure 5: Tensor-DTI architecture with pocket embeddings. The model extends the base architecture by incorporating binding pocket representations, enabling site-specific interaction modeling. The protein shown is 5ISX. The SaProt image is adapted from (Su et al., 2023), and the GearNet image is adapted from (Zhang et al., 2023b).

B LEARNING POCKET REPRESENTATIONS

Our model combines the strengths of protein language models and geometric learning. Evolutionary embeddings from ESM-2 capture functional conservation, while GearNet processes residue-level spatial interactions. The model operates as follows:

B.1 SEQUENCE ENCODING

ESM-2 generates residue embeddings via transformer layers. Each residue is transformed into a feature vector:

$$h_i^{(0)} = \text{Embedding}(r_i) \in \mathbb{R}^d, \tag{1}$$

where d is the embedding dimension. The embeddings are refined through multiple transformer layers:

$$\alpha_{ij}^{(l)} = \text{Softmax}j\left(\frac{W_q h_i^{(l)\top} W_k h_j^{(l)}}{\sqrt{d}}\right) \tag{2}$$

$$h_i^{(l+0.5)} = h_i^{(l)} + \sum j = 1^n \alpha_{ij}^{(l)} W_v h_j^{(l)} \tag{3}$$

$$h_i^{(L)} = \text{LayerNorm}(h_i^{(l+0.5)}) \tag{4}$$

B.2 STRUCTURAL GRAPH REPRESENTATION

GearNet constructs a residue graph, where nodes represent residues and edges define relationships. Each node is initialized with its corresponding ESM-2 sequence embedding:

$$u_i^{(0)} = h_i^{(L)} \tag{5}$$

The structural information is integrated through message-passing layers:

$$u_i^{(l+1)} = u_i^{(l)} + \sigma\left(\sum_{r\in\mathcal{R}} \sum_{j\in\mathcal{N}_r(i)} W_r u_j^{(l)}\right) \tag{6}$$

where $\mathcal{N}_r(i)$ represents neighbors of node i connected by edge type r.

B.3 PREDICTION ARCHITECTURE

The combined embeddings are processed through a multilayer perceptron (MLP) classifier:

$$h_i = \text{GELU}(W_1 x_i + b_1) \tag{7}$$

$$y_i = W_2 h_i + b_2 \tag{8}$$

where y_i represents the predicted binding probability for residue i.

B.4 POCKET EXTRACTION

High-confidence residues (binding probability > 0.5) are clustered using DBSCAN, with distance threshold $\epsilon = 5$Å and minimum cluster size 3.

B.5 PERFORMANCE AND GENERALIZATION

We evaluated our model on the LIGYSIS benchmark, achieving the highest F1 score (0.42) and competitive MCC (0.37), surpassing established methods such as P2Rank, PUResNet, and GrASP. Our model also excels in cryptic pocket detection, outperforming PocketMiner on apo structures without explicit cryptic pocket training.

C Hyperparameter Configurations and Model Architectures

C.1 DTI

We trained the DTI model with a combination of contrastive loss and binary cross-entropy (BCE) loss. The contrastive loss encourages positive drug–target pairs to cluster together in latent space while pushing negative pairs apart. Specifically, for a positive pair (d, p) and corresponding negative pairs, we minimized:

$$L_{\text{contrastive}} = \sum_{(d,p)} \max\Big(0, \alpha + \|f_d(h_G) - f_p(h_P)\|_2 - \|f_d(h_G) - f_p(h_P^-)\|_2\Big), \tag{9}$$

where α is a margin hyperparameter and h_P^- denotes a non-interacting protein embedding. The BCE loss is:

$$L_{\text{BCE}} = -\sum_{(d,p)} [y_{dp} \log(\hat{y}_{dp}) + (1 - y_{dp}) \log(1 - \hat{y}_{dp})], \tag{10}$$

where y_{dp} is the ground-truth interaction label. We combined these losses to achieve robust, discriminative embeddings suited for both classification and interpretability.

We optimized the parameters using the Adam optimizer with a learning rate of 5×10^{-5} for DTI tasks, weight decay of 1×10^{-5}, and early stopping based on validation performance. Multiple runs ensured statistical robustness, and final reported metrics were averaged across runs.

Tensor-DTI integrates pocket embeddings to refine binding-site specificity. These embeddings, derived using ESM-GearNet (Zhang et al., 2023a), capture residue-residue interactions within functional binding sites. To combine protein and pocket representations, we apply a weighted aggregation:

$$\text{combined_protein_pocket} = \alpha \cdot \text{encoded_protein} + \beta \cdot \text{encoded_pocket} \tag{11}$$

For all results reported in this study, we set $\alpha = 1$ and $\beta = 2$. This weighting emphasizes the binding site information while retaining global protein context.

Table 3: Hyperparameters used for DTI models in protein-drug interaction benchmarks.

Benchmark	**Emb. Dim**	**Hidden Dim**	**Output Dim**
BIOSNAP, BindingDB, DAVIS	(64, 1280)	512	256
DUD-E	(64, 1280)	512	256
SMPBind	(64, 1280)	1024	512
PLINDER (No Pocket)	(64, 1280)	512	256

Benchmark	**LR**	**Epochs**
BIOSNAP, BindingDB, DAVIS	0.00005	1000
DUD-E	0.000005	300
SMPBind	0.00001	100
PLINDER (No Pocket)	0.00001	1000

Table 4: Hyperparameters used for DTI models in alternative biomolecular interaction benchmarks, including RNA, peptides, and pocket embeddings.

Benchmark	**Emb. Dim**	**Hid. Dim**	**Out. Dim**
PLINDER (With Pocket)	(64, 1280, 1536)	512	256
CoPRA (Protein-RNA)	(480, 1280)	512	256
Propedia (Peptide-Protein)	(480, 1280)	512	256

Benchmark	**LR**	**Epochs**
PLINDER (With Pocket)	0.00001	1000
CoPRA (Protein-RNA)	0.00001	1000
Propedia (Peptide-Protein)	0.00001	1000

C.2 DTA

Table 5: Hyperparameters used for DTA models.

Benchmark	**Emb. Dim**	**Hidden Dim**	**Output Dim**
TDC-DG (Molecule-Protein)	(2048, 1280)	4096	1024
LP-PDBBind (Leakproof)	(64, 1280)	4096	1024
PDBBind (Molecule-Protein)	(64, 1280)	4096	1024
PDBBind (Peptide-Protein)	(480, 1280)	4096	1024
PDBBind (RNA-Drug)	(480, 64)	4096	1024

Benchmark	**LR**	**Epochs**
TDC-DG (Molecule-Protein)	0.0001	1000
LP-PDBBind (Leakproof)	0.0001	1000
PDBBind (Molecule-Protein)	0.0001	1000
PDBBind (Peptide-Protein)	0.0001	1000
PDBBind (RNA-Drug)	0.0001	200

D CONFIDENCE MODEL

To enhance prediction robustness and reliability, we leveraged the largest available dataset (SMPBind - I) for training. This ensured the model was exposed to a diverse range of chemical configurations, improving its ability to generalize and accurately identify novel drug-target interactions with high confidence.

To evaluate the reliability of its predictions, we introduce a Confidence Model, which is trained jointly with the primary model. The model processes the drug-target embeddings and interaction logits, estimating the certainty of each prediction through a continuous confidence score.

The confidence model is implemented as a feedforward neural network, which takes as input the concatenated drug-target embeddings and interaction logits, producing a single confidence score:

$$c = f_{\text{conf}}(E_{\text{combined}}, \hat{y}),$$

where E_{combined} represents the fused representation of the drug-target pair, and $\hat{y}$ is the predicted interaction score.

The confidence score is designed to estimate the absolute difference between the ground-truth label y (which is either 0 or 1) and the predicted interaction score $\hat{y}$. This means that the closer the confidence score is to 0, the more reliable the prediction is, while values closer to 1 indicate higher uncertainty or potential misclassification.

Training is performed jointly with the primary model using a combined loss function. To align the confidence scores with prediction reliability, the confidence model is optimized using mean squared error (MSE) loss. This loss function ensures that the confidence score directly reflects the prediction's deviation from the ground truth, reinforcing the model's ability to assess uncertainty.

$$L_{\text{Conf}} = \frac{1}{N} \sum_{i} \left(c_i - |y_i - \hat{y}_i|\right)^2 .$$

The final loss function integrates confidence estimation with interaction prediction:

$$L_{\text{Final}} = \alpha L_{\text{Prediction}} + (1 - \alpha) L_{\text{Conf}},$$

where $0 \leq \alpha \leq 1$ controls the trade-off between classification performance and confidence calibration.

After training, the model assigns confidence scores to predictions, providing a measure of reliability. To better understand how confidence correlates with prediction accuracy, we categorize predictions into True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN).

Figure 6 illustrates the distribution of confidence scores across these categories for a test subset of SMPBind - I. As expected, correctly classified samples (TP and TN) exhibit lower confidence scores, indicating high certainty in their predictions. In contrast, misclassified samples (FP and FN) tend to have higher confidence scores, reflecting greater uncertainty.

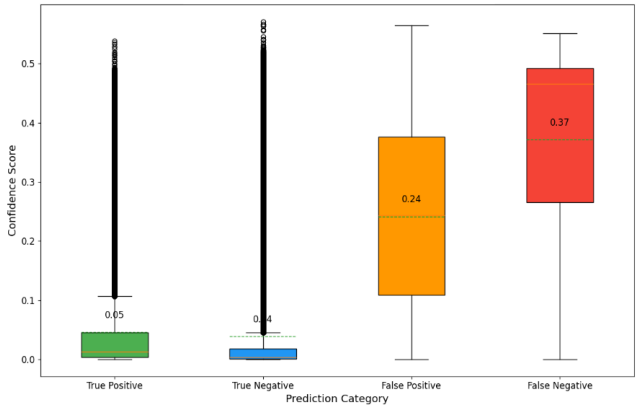


Figure 6: Distribution of confidence scores across prediction categories (TP, FP, TN, FN). Lower confidence scores indicate higher certainty in the model's predictions, while higher scores suggest increased uncertainty.

This confidence-aware framework enhances the interpretability of predictions, providing a systematic way to prioritize high-confidence interactions for downstream applications. Moreover, it serves as the method used to rank predictions, enabling the selection of interacting drug-protein pairs from a vast pool of possible combinations. By sorting based on confidence scores, the approach ensures that the top-ranked interactions are not only predicted as interacting but also assigned a high confidence (i.e., a lower confidence score), making them strong candidates for further investigation.

E Ablation Studies on Embedding Effectiveness

E.1 Ablation Studies for DTI

We conducted extensive ablation experiments on widely used DTI benchmarks—BIOSNAP, BindingDB, and DAVIS—to identify the most effective drug and protein embeddings for accurate interaction prediction. Below, we present our findings and the rationale for the selected configurations.

E.1.1 Drug Embeddings

We evaluated multiple drug embedding strategies to determine the most effective representation for our model:

- **Mollm**: A transformer-based model that extracts molecular features through self-attention mechanisms (Vaswani et al., 2017; Radford et al., 2021).
- **GIN with Barlow Twins Loss**: A self-supervised method using Graph Isomorphism Networks to learn robust molecular representations (Zbontar et al., 2021).
- **Cooperative Protocol (COOP)**: An embedding approach integrating cooperative strategies to improve drug representations (Finkelshtein et al., 2023).
- **Molecular FingerPrints (MFPS)**: Provides robust and detailed encodings of molecular structures by converting molecules into fixed-length binary vectors representing the presence or absence of particular substructures (Rogers & Hahn, 2010).
- **Graph Convolutional Network (GCN)**: A neural network model that effectively captures the structural features of drugs by transforming molecular graphs into high-dimensional embeddings suitable for interaction prediction (Kipf & Welling, 2016).

For the initial ablation study, we used ESM-2 for protein embeddings and tested different drug embeddings. The results are presented in Table 6.

Table 6: Performance of Tensor-DTI on different datasets with various drug embeddings. GIN_L differs from GIN by using a larger training dataset and a more complex architecture. All values correspond to AUPR.

Dataset/d. emb.	GCN	GIN_L	MFPS	GIN	COOP
BIOSNAP	<u>0.879</u>	0.832	**0.881**	0.837	0.837
unseen T	<u>0.708</u>	0.646	**0.720**	0.649	0.638
unseen D	**0.872**	0.827	<u>0.851</u>	0.832	0.832
BindingDB	<u>0.664</u>	0.583	**0.679**	0.591	0.581
DAVIS	**0.532**	0.331	<u>0.527</u>	0.334	0.338

GCN performed particularly well on unseen drugs, while MFPS achieved the highest overall AUPR scores across benchmarks. These findings confirm their robustness in different settings, leading to their selection for further analysis.

E.1.2 Protein Embeddings

After evaluating drug embeddings, we assessed the impact of protein embeddings, comparing SaProt and ESM-2 to determine their effect on model performance.

- **SaProt Embeddings**: Derived from a transformer-based model specifically designed for protein sequences, offering high-quality embeddings (Su et al., 2023).
- **ESM-2 Embeddings**: Generated by a state-of-the-art transformer model trained on a large corpus of protein sequences, known for its robust performance (Lin et al., 2023).

The results of this ablation study are shown in Table 7.

Table 7: Performance of Tensor-DTI on different datasets with various protein embeddings.

Dataset/d. emb.	GCN (ESM-2)	MFPS (ESM-2)	GCN (SaProt)	MFPS (SaProt)
BIOSNAP	0.879	0.881	**0.897**	0.894
unseen T	0.708	0.720	0.836	**0.838**
unseen D	0.872	0.851	**0.879**	0.849
BindingDB	0.664	0.679	0.685	**0.689**
DAVIS	0.532	0.527	**0.555**	0.552
MEAN	0.731	0.732	**0.770**	0.764

The results indicated that SaProt embeddings consistently outperformed ESM-2 embeddings across multiple datasets, leading us to select SaProt for protein embeddings in our final model. We also selected GCN as the technique for further analysis.

E.1.3 Impact of Training the GCN

To assess the impact of large-scale training, we examined whether pretraining the GCN on a larger dataset (PCBA_1328) improves performance. As shown in Table 8, the pretrained GCN consistently outperforms its untrained counterpart, achieving higher AUPR scores across all benchmarks.

Table 8: Performance of Tensor-DTI with GCN Trained embeddings and ConPlex.

Dataset/d. emb.	GCN	**GCN trained**	ConPlex(MFPS)
BIOSNAP	0.900 ± 0.002	**0.903 ± 0.003**	0.897 ± 0.001
unseen T	0.834 ± 0.004	0.839 ± 0.003	**0.842 ± 0.006**
unseen D	0.880 ± 0.004	**0.888 ± 0.002**	0.874 ± 0.002
BindingDB	0.686 ± 0.003	**0.699 ± 0.002**	0.628 ± 0.012
DAVIS	0.544 ± 0.015	**0.547 ± 0.006**	0.458 ± 0.016

These findings demonstrate that both GCN and trained GCN embeddings significantly outperform ConPlex in most benchmarks, reinforcing the robustness of our proposed Tensor-DTI model.

E.2 Ablation Studies for DTA

In addition to our comprehensive analysis for DTI, we conducted ablation studies for DTA prediction in the TDC-DG benchmark to identify the optimal embeddings for both drugs and proteins.

E.2.1 Protein Embeddings

We first compared the performance of different protein embeddings, specifically SaProt and ESM-2 embeddings:

- **SaProt Embeddings** (Su et al., 2023).
- **ESM-2 Embeddings** (Lin et al., 2023).

The results of this ablation study are shown in Table 9.

Table 9: Performance of Tensor-DTI in terms of PCC on TDC-DG with two different protein embeddings.

t. emb./d. emb.	GCN
ESM-2	**0.550**
SaProt - 650M	0.530

Based on these results, ESM-2 embeddings were selected due to their superior performance.

E.2.2 DRUG EMBEDDINGS

We then evaluated two primary drug embedding methods for our DTA model:

- **Molecular FingerPrints (MFPS)** (Rogers & Hahn, 2010).
- **Graph Convolutional Network (GCN)** (Kipf & Welling, 2016).

For this ablation study, we used ESM-2 for protein embeddings and evaluated different drug embeddings. The results are presented in Table 10.

Table 10: Performance of Tensor-DTI and ConPlex on different datasets with various drug embeddings (PCC).

-/d. emb.	GCN	MFPS	GCN trained	ConPlex (MFPS)
PCC	0.546 ± 0.02	**0.580 ± 0.004**	0.539 ± 0.001	0.538 ± 0.008

Among the evaluated methods, MFPS achieved the highest PCC score (0.580), demonstrating superior performance compared to other embedding strategies. Given its consistently strong results across benchmarks, we selected MFPS as the preferred drug embedding for the final model.

Additionally, Table 10 confirms that in this case, training the GCN did not provide any advantage for the DTA study. This may be attributed to the activity information contained in the PCBA_1328 dataset.

E.3 ABLATION STUDIES FOR DTA - LEAK PROOF BENCHMARK

To further assess the impact of embedding choices on DTA prediction under strict data leakage constraints, we evaluated Tensor-DTI on the LP-PDBBind benchmark. The results highlight significant differences in performance based on the selected embeddings.

The best-performing configuration combined SaProt protein embeddings with trained GCN drug embeddings, achieving a PCC of 0.565 and an RMSE of 1.62. In contrast, using ESM-2 protein embeddings with MFPS drug embeddings led to a lower PCC of 0.450 and a higher RMSE of 1.79. Overall, the best-performing configuration combined SaProt protein embeddings with trained GCN drug embeddings, achieving the highest PCC and lowest RMSE. These findings highlight the importance of structural protein embeddings (SaProt) and graph-based drug representations (trained GCN) in improving model generalization under strict data leakage constraints. This underscores the critical role of domain-specific, structure-informed embeddings in achieving robust and accurate affinity predictions in real-world applications.

F DATABASES

Data collection, processing and splitting are pivotal in drug-target interaction predictions. We outline all datasets used in this work with detailed descriptions on the train-validation-test splittings performed over them.

BIOSNAP. This drug-target interaction network provides information on the genes (i.e., proteins encoded by genes) targeted by drugs available on the U.S. market. Drug targets are molecules essential for the transport, delivery, or activation of a drug. This information is widely utilized in computational drug target discovery, drug design, docking or screening, metabolism prediction, interaction prediction, and general pharmaceutical research. Drug target information includes biotech drugs and nutraceuticals. On average, these drugs have 5–10 unique target proteins. The dataset lists all known targets with physiological or pharmaceutical effects, not just a single primary target, and fully accounts for the fact that many drug targets are protein complexes composed of multiple subunits or combinations of proteins.

Preprocessing and splitting. We use ChGMiner from BIOSNAP which contains only positive drug-target interactions. Following the approach described in (Singh et al., 2023), we create negative DTIs

by randomly sampling an equal number of protein–drug pairs, under the assumption that a random pair is unlikely to interact positively.

DAVIS. The DAVIS dataset is a comprehensive resource profiling interactions between 72 kinase inhibitors and 442 kinases, covering over 80% of the human catalytic protein kinome. It provides detailed binding affinity data (Kd) for each interaction and calculates selectivity scores to evaluate inhibitor specificity. The dataset distinguishes between type I inhibitors, which target active kinase conformations, and type II inhibitors, which bind inactive states, showing that type II inhibitors are generally more selective, though exceptions exist. It highlights group-selective inhibitors, off-target profiles, and structural features contributing to selectivity, making it invaluable for drug discovery, kinase biology, and computational modeling. This resource has facilitated the development of selective inhibitors and the repurposing of existing drugs, while offering insights into kinase-inhibitor interaction patterns and signaling pathways.

BindingDB. BindingDB is a publicly accessible database that provides experimentally determined protein-ligand binding affinities, focusing on drug-target interactions. It currently contains over 20,000 binding measurements for approximately 11,000 small molecule ligands and 110 protein targets, including isoforms and mutants. BindingDB integrates data from enzyme inhibition studies and isothermal titration calorimetry, extracted from scientific literature and directly deposited by experimentalists. The database is designed to support diverse applications, such as computational drug design, ligand discovery, and structure-activity relationship analysis. Its web interface offers powerful tools for querying by chemical structure, substructure, protein sequence, or affinity ranges, and supports virtual screening using uploaded compound databases. By linking data to the Protein Data Bank (PDB) and PubMed, BindingDB facilitates the integration of binding, structural, and sequence data, making it a valuable resource for researchers in pharmaceutical sciences and computational biology.

Preprocessing and splitting for DAVIS and BindingDB datasets. Following the approach described in (Singh et al., 2023), we treat pairs with $K_D < 30$ as positive DTIs and those with larger K_D values as negative DTIs. The dataset is split into 70% for training, 10% for validation, and 20% for testing. Training data are subsampled to have an equal number of positive and negative interactions, ensuring a balanced training set, while validation and test data retain the original ratio of interactions. This preprocessing strategy ensures consistency across datasets and facilitates robust model evaluation. Compared to DAVIS, which represents a low-data learning setting with 2,086 training interactions, BindingDB provides a broader learning scenario with 12,668 training interactions, offering greater diversity in drug-target interaction pairs. Both datasets complement each other, enabling evaluation of model performance across varying levels of data availability.

DUD-E (Kinase Subset). DUD-E provides a curated set of protein targets and their known binding partners, along with decoy molecules designed to resemble the physicochemical properties of true binders but are experimentally confirmed not to bind. From this database, we focus specifically on the kinase family, which consists of 26 protein targets. For each target, the dataset includes an average of 224 true binding partners and 50 decoys per binding partner, enabling robust evaluation of model performance in distinguishing between true and decoy interactions.

Preprocessing and splitting. We derive train-test splits by partitioning the kinase targets such that no target appears in both the training and test sets. Specifically, we hold out 50% of kinase targets for testing and use the remaining targets for training, ensuring representative coverage of the kinase family in both splits. This setup evaluates the ability of models to generalize to unseen kinase targets and to distinguish true binding interactions from decoy molecules.

PLINDER. PLINDER (Durairaj et al., 2024) is a comprehensive protein-ligand interaction dataset designed to address critical challenges in computational drug design and protein engineering. It includes 449,383 systems, each extensively annotated with over 500 attributes, including protein, ligand, pocket, and interaction-level similarity metrics. PLINDER uniquely links holo systems to apo and predicted structures, enabling realistic evaluation scenarios such as docking, ligand generation, and co-folding. By employing advanced splitting algorithms, PLINDER minimizes information leakage, enhancing the evaluation of machine learning models' generalization capabilities. Its rich annotations, task-specific test sets, and robust evaluation frameworks make PLINDER a valuable resource for advancing predictive modeling in protein-ligand interactions.

Preprocessing and splitting. PLINDER preprocesses and splits its extensive protein-ligand interaction (PLI) dataset, comprising 449,383 systems from 162,978 PDB entries, into training, validation, and test sets with rigorous quality control and annotation. A total of 113,498 high-quality systems meeting stringent criteria such as resolution ≤ 3.5 Å and R-factor ≤ 0.4 are annotated with over 500 features, including protein, pocket, ligand, and interaction-level details. Train-test splits are generated using graph-based algorithms that classify systems based on similarity metrics like protein sequence identity, pocket overlap, and ligand fingerprints, ensuring minimal leakage and maximum diversity. The PLINDER-PL50 split includes 57,602 training, 3,453 validation, and 3,729 test systems, achieving 0% leakage for PLI similarity $\geq 50\%$. Another configuration, PLINDER-ECOD, defines splits using evolutionary domains and comprises 77,411 training, 10,169 validation, and 12,174 test systems, all containing 100% high-quality systems to support robust and realistic benchmarking of computational models.

SMPBind - I. The following dataset is a curated mix of several databases. These databases include ChEMBL, PubChem, ChEBI (Chemical Entities of Biological Interest), STITCH, OpenTargets, DGIdb (Drug Gene Interaction Database), Pharos, TTD (Therapeutic Targets Database), HMDB (Human Metabolome Database), T3DB (Toxin and Toxin-Target Database), BindingDB and DTC (Drug Target Commons).

Preprocessing and splitting. From these databases we extract the pairs of protein molecules that have been experimentally validated at least once. Afterwards, we realize a heavy de-duplication procedure. Racemic mixtures are separated into their chiral parts, hydrogen atoms are removed, metal atoms are disconnected from the molecule, the molecule is normalized and reionized. After this point, in the case that the molecule has several fragments, the biggest one is assumed to be the bioactive one, so it is selected. Then the molecule is neutralized and canonized, to avoid the presence of tautomerism overlap within the database. Lastly, InchiKeys are computed from the resulting molecules and used for de-duplication. The resulting database contains more than 400,000 different Murckro scaffolds, and more than 35,000 unique proteins, divided in over 7,000 different families and 1,000 superfamilies.

Propedia. Propedia v2.3 (Martins et al., 2023) is a peptide-protein interaction database. The last updated version builds on the foundational Propedia database by incorporating over 49,300 peptide-protein complexes—a 150% increase from its initial release—and introducing graph-based structural signatures to represent peptide structures numerically. These signatures, calculated using the aCSM-ALL algorithm, enhance the ability to cluster and analyze peptides based on sequence similarities, structural interfaces, and binding sites. Propedia v2.3 supports machine learning applications, offering a CSV dataset of feature vectors suitable for tasks like peptide classification and therapeutic discovery. The database facilitates in-depth exploration of peptide-protein recognition mechanisms, a critical aspect of drug development and biotechnology.

Preprocessing and splitting. We preprocess the Propedia dataset by clustering protein and peptide embeddings into distinct groups using K-Means. Protein and peptide embeddings are numerically represented. Clustering ensures that similar protein and peptide structures are grouped together, facilitating representative splitting across training, validation, and test sets. The dataset is split into 80% for training, 10% for validation, and 10% for testing, ensuring that no peptide-protein pairs from the same cluster appear across different splits. Negative pairs are generated by randomly sampling an equal number of peptide-protein pairs within each split, resulting in a 1:1 balance of positive and negative interactions in all sets. The original dataset contains only positive pairs, and this augmentation ensures balanced training and evaluation.

CoPRA dataset. The CoPRA dataset (Han et al., 2024) is designed for protein-RNA binding affinity prediction and consists of two subsets: PRI30k (training), and PRA310 (test). This dataset is directly provided by CoPRA and includes 30,000 non-redundant protein-RNA complexes from BioLiP2 (PRI30k) and 310 high-quality complexes curated from PDBBind, PRBABv2, and ProNAB (PRA310).

Preprocessing and splitting. Positive interactions are extracted from experimental annotations, while negative pairs are generated by random pairing within each subset to ensure a 1:1 positive-negative ratio. Clusters are created using CD-HIT at 70% sequence identity to prevent data leakage, with distinct splits for training, validation, and testing. This setup ensures diverse, high-quality data for robust model evaluation (Han et al., 2024).

G Overview of Dataset Volumes Across Benchmarks

Table 11: Dataset volumes for DTI benchmarks across train, validation, and test splits, showing the number of positive and negative pairs for each interaction type. Note that for the DUD-E dataset, no validation set was used.

Target	Ligand	Benchmark	Train		Validation		Test	
			Positive	Negative	Positive	Negative	Positive	Negative
Protein	Drug	BIOSNAP	9,490	9,306	1,372	1,327	2,718	2,656
		BINDINGDB	5,842	5,702	859	5,134	1,752	10,307
		DAVIS	883	909	132	2,474	252	4,987
		DUD-E - Kinase	4,112	150,027	-	-	5,027	201,599
		PLINDER	692,144	692,129	32,252	32,248	49,770	49,760
		PLINDER pocket	311,978	388,522	12,730	22,243	15,577	23,754
		SMPBind	10,340,470	10,349,292	1,292,375	1,293,821	1,292,591	1,293,639
Protein	Peptide	Propedia	40,370	40,393	4,418	4,419	4,415	4,412
Protein	RNA	CoPRA	15,626	15,626	820	820	200	200

Table 12: Dataset volumes for DTA benchmarks, showing the number of positive pairs for train, validation, and test splits for each interaction type.

Target	Ligand	Benchmark	Train (Positive)	Validation (Positive)	Test (Positive)
Protein	Drug	TDC-DG	146,744	36,686	49,028
		LP-PDBBIND	5,691	1,317	3,103
		LP-PDBBIND (ΔG)	5,691	1,317	3,103
		PDBBind-Opt	13,185	1,465	1,628
		PDBBind-Opt+LP	7,051	1,846	4,193
Protein	Peptide	PDBBind-Opt	1,896	210	240
RNA	Drug	PDBBIND	96	13	11
Protein	RNA	CoPRA	165	21	14

H Low-Leakage Datasets Results

To assess Tensor-DTI's robustness in minimized leakage scenarios, we compare its performance against a one-hot encoding baseline across multiple datasets, ensuring consistency in hyperparameter settings.

Table 13: Performance comparison of DTI (classification) and DTA (regression) models on minimized leakage datasets. AUPR is used for classification benchmarks, while PCC and RMSE evaluate affinity prediction tasks.

Benchmark	**Model**	**AUPR**	**PCC**	**RMSE**
PLINDER (random negatives)	Tensor-DTI	0.779 ± 0.001	-	-
PLINDER (high pocket dissimilarity)	Tensor-DTI	0.698 ± 0.005	-	-
LP-PDBBind				
	Tensor-DTI	-	0.565 ± 0.004	1.620 ± 0.024
	DeepDTA	-	0.512 ± 0.020	2.290 ± 0.040
	AutoDock Vina	-	0.450 ± 0.000	2.560 ± 0.000
	One-Hot	-	0.428 ± 0.016	2.287 ± 0.032
LP-PDBBind (ΔG prediction)				
	Tensor-DTI	-	0.528 ± 0.013	2.122 ± 0.032
	One-Hot	-	0.428 ± 0.016	2.287 ± 0.032
PDBBind-Opt Peptide-Protein				
	Tensor-DTI	-	0.679 ± 0.014	1.175 ± 0.020
	One-Hot	-	0.568 ± 0.025	1.846 ± 0.099
PDBBind-Opt Small Molecule-Protein				
	Tensor-DTI	-	0.750 ± 0.005	1.335 ± 0.011
	One-Hot	-	0.728 ± 0.007	1.320 ± 0.012
PDBBind-Opt Small Molecule-Protein (Leak proof split)	Tensor-DTI	-	0.493 ± 0.005	1.545 ± 0.006
	One-Hot	-	0.385 ± 0.014	1.752 ± 0.033

I EXPANDED RESULTS ON BIOMOLECULAR INTERACTION PREDICTIONS

We compare Tensor-DTI against a one-hot encoding baseline under the same hyperparameter settings to evaluate its performance across various biomolecular interaction tasks.

Table 14: Performance of Tensor-DTI on the Propedia peptide interaction database.

Model	**AUPR**
Tensor-DTI	0.953 ± 0.001
One-hot encoding	0.884 ± 0.003

Table 15: Performance of Tensor-DTI on CoPRA.

Model	**AUPR**
Tensor-DTI	0.916 ± 0.008
One-hot encoding	0.795 ± 0.009

Table 16: Performance of Tensor-DTI on the CoPRA (PRA310). The top table corresponds to K_d (binding constant) prediction, while the bottom table corresponds to ΔG (free energy) prediction.

K_d Prediction

Model	**PCC**	**RMSE**
Tensor-DTI	0.631 ± 0.111	1.443 ± 0.232
One-hot encoding	0.468 ± 0.189	1.399 ± 0.232

ΔG Prediction

Model	**PCC**	**RMSE**
Tensor-DTI	0.621 ± 0.052	1.910 ± 0.212
One-hot encoding	0.453 ± 0.213	1.896 ± 0.430

Table 17: Performance of Tensor-DTI on the PDBBind interaction database, selecting from it the Drug-RNA interactions.

Model	**PCC**	**RMSE**
Tensor-DTI	0.792 ± 0.015	1.684 ± 0.038
One-hot encoding	0.633 ± 0.018	1.738 ± 0.036