OXFORD

Structural bioinformatics

GS-DTI: a graph-structure-aware framework leveraging large language models for drug–target interaction prediction

Qinze Yu[1],[†], Chang Zhou[1],[†], Jiyue Jiang[1], Xiangyu Shi[2], Yu Li[1],[3],∗ 🄳

[1]Department of Computer Science and Engineering, CUHK, Hong Kong SAR 999077, China
[2]Department of Computer Science and Technology, Beijing JiaoTong University, Beijing 100044, China
[3]The CUHK Shenzhen Research Institute, Shenzhen 518057, China

†= equal contribution.

*Corresponding author. Department of Computer Science and Engineering, CUHK, Hong Kong SAR 999077, China; The CUHK Shenzhen Research Institute, Shenzhen 518057, China. E-mail: liyu@cse.cuhk.edu.hk.

Associate Editor: Arne Elofsson

Abstract

Motivation: Accurate and generalizable prediction of drug–target interactions (DTIs) remains a critical challenge for drug discovery, particularly when addressing underexplored targets and compounds. Recent advances in graph neural networks and large-scale pre-trained models offer new opportunities to capture rich structural and functional features essential for DTI prediction while enhancing the generalization ability.

Results: We present GS-DTI, a graph structure-based DTI prediction framework that integrates molecular graph transformers, protein language models, and protein tertiary structure. Our method achieved robust and interpretable DTI predictions. GS-DTI extracts drug features from SMILES-derived molecular graphs using a knowledge-guided pre-trained transformer, while protein features are derived from both sequence and predicted 3D structure for comprehensive representation. A multi-task loss function equipped with contrastive learning is adopted to enhance generalization and functional interpretability. Extensive experiments on the benchmarks and challenging cross-domain settings demonstrate that GS-DTI achieves state-of-the-art performance. Notably, our model improves the MCC by over 10% compared to previous methods in the drug–target pair cold start test. The model can pinpoint the binding pockets of the targets, offering robust interpretability, and case studies show GS-DTI's promising potential in virtual screening for new candidate drugs of BACE1.

Availability and implementation: The GS-DTI source code and processed datasets are available at https://github.com/purvavideha/GSDTI. All experimental data are derived from public sources.

1 Introduction

Drug–target interaction (DTI) prediction is of great importance in the process of drug discovery (Chen *et al.* 2016, Huang *et al.* 2021b), as drugs exert their effects primarily by binding to specific protein targets. A drug is usually a chemical compound that induces physiological changes in the body when consumed, injected, or absorbed. A target, which is normally a protein, can be recognized or bound by substances such as ligands or drugs, enabling it to interact with or be influenced by these molecules. Traditional experimental methods for the measurement of drug–target interaction affinity are time- and cost-consuming, preventing analysis of data at scale (Deng *et al.* 2022). To improve efficiency, some in silico approaches were proposed. One class is docking simulations, which use the 3D structure of drugs and targets to identify their potential binding sites and binding affinity (Li *et al.* 2006, Pinzi and Rastelli 2019). However, these methods heavily rely on accurate 3D structure data, which may not be available for all samples, and they are based on well-defined binding sites, hence have difficulty detecting unknown sites. Second, there are similarity-based and network-based

methods, which use protein-protein similarity, and drug–drug similarity to make inferences informed by the known DTI (Ding *et al.* 2014).

Similarity-based and network-based methods are not able to generalize to data that are absent from the training set. With the advancement of statistical and computational methodologies, machine learning frameworks were developed to resolve this problem. They achieved decent accuracy and enabled large-scale data processing in an acceptable time. The normal form of these methods is to extract the drug, target, and their interaction information and then systematically integrate them to build predictive models, which are dedicated to binary classification or regression tasks. For instance, DeepDTA (Öztürk *et al.* 2018), 3DProtDTA (Voitsitskyi *et al.* 2023), and DeepConv-DTI (Lee *et al.* 2019) extract molecular features of drugs and targets with Convolutional Neural Networks (CNN) and combine them for prediction. Considering the structures of drugs and targets are intrinsically graph-like, and the relationship between drugs and targets can be modeled with graphs, a wide range of Graph Neural Networks (GNN) based methods have been proposed. GraphDTA (Nguyen *et al.* 2021) and

DrugBAN (Bai *et al.* 2023) take drug compounds as a graph of the interactions between atoms and utilize GNN for representation learning. ZeroBind (Wang *et al.* 2023) builds the graph of proteins based on the distance between amino acids in space. MINDG (Yang *et al.* 2024) constructs drug–target relationship graphs and uses Graph Attention Networks (GAT) for feature aggregation. Then, due to the improving availability of computational resources, molecule pre-trained models (Li *et al.* 2022) and protein language models (Elnaggar *et al.* 2022, Lin *et al.* 2023) are developed and show impressive performance on various downstream tasks (Yu *et al.* 2021, Shen *et al.* 2024). DLM-DTI (Lee *et al.* 2024) takes both molecule and protein pre-trained models for feature enrichment.

However, there are still some challenges with existing computational approaches for DTI prediction, which limit their effectiveness in real-world scenarios. First and foremost, they do not demonstrate robustness in domain-shift data prediction and may struggle to predict interactions involving rare drugs or targets that have low similarity to known data. The distribution of the training data can greatly vary from the distribution of the data in testing or actual applications. Therefore, the lack of generalizability can hinder the performance in identifying interactions for new or underexplored drug–target pairs. Furthermore, most methods lack interpretability, making it difficult to understand how the models make inferences. This black-box nature limits their utility in drug development, where understanding the molecular basis of drug–target interactions is crucial for designing effective therapies. Without clear explanations of DTI mechanisms, the model may not be considered trustworthy.

In this study, we addressed the above problems by leveraging the molecule graph and protein language models as well as the protein tertiary structures, and proposed a Graph Structure-based Drug Target Interaction prediction method (GS-DTI) for accurate DTI identification. Specifically, our model takes the Simplified Molecular Input Line Entry System (SMILES) of drugs and the protein amino acid sequences as input. For drugs, the SMILES strings are first converted into molecule graphs, and then a pre-trained molecular graph transformer (KPGT) (Li *et al.* 2022) is applied to represent embeddings of drug graphs. For targets, we used ESMFold (Lin *et al.* 2023) to predict the protein tertiary structures and obtain contact maps, and ESM-2 (Lin *et al.* 2023) to generate per-amino acid embeddings. We then built the protein graphs that use contact maps as edge information and amino acid embeddings as node features. After that, Graph Multiset Transformer (GMT) (Baek *et al.* 2021, Gu *et al.* 2023), a multi-head attention-based global pooling layer for capturing the interaction between nodes, is applied for protein message aggregation. The drug and target features are combined for DTI prediction. Here, we also adopted contrastive learning to enhance the representation quality and generalization of the drug and target.

We compared GS-DTI with several DTI prediction deep learning methods on both in-domain and cross-domain settings. The results demonstrated that our approach surpasses the other methods, achieving superior overall performance. Particularly, our model achieves more than 10% improvement of MCC over previous methods on the drug–target pair cold start test. Then, by checking the contribution of each amino acid for prediction, we found that GS-DTI could help detect the target binding pocket, providing interpretable insights for the prediction results. At last, we used GS-DTI to successfully identify effective inhibitors of BACE1, an unseen target to our model, indicating our model is a generalized and useful tool for discovering novel drugs for targets. These results suggest that GS-DTI can effectively predict DTI and greatly facilitate drug discovery.

2 Materials and methods

2.1 Datasets

2.1.1 Training data

BindingDB (Liu *et al.* 2007) is a public, web-accessible database that provides information about experimentally measured binding affinities, primarily focusing on the interactions between proteins (commonly regarded as drug targets) and small, drug-like molecules. We access the data through the Therapeutics Data Commons (TDC) (Huang *et al.* 2021a). We used interactions in the BindingDB dataset that are measured in K_d for training. The data labels (affinity values) are first converted into pK_d (log scale), and then we set the threshold as 7 to binarize them. Considering our computational resource limitations preclude structure prediction for proteins with particularly long sequences, we removed all interaction data whose target lengths are longer than 1000, which is approximately the maximum length that we can handle. The processed dataset contains 33,238 negative samples and 8,846 positive samples, including 9,833 unique drugs and 1,235 unique proteins (Table 1, available as supplementary data at *Bioinformatics* online).

2.1.2 Test data

Two datasets, named Davis and BIOSNAP, are used for the independent tests. Data in the Davis set are derived from experimental binding affinity measurements between kinase inhibitors and kinases. The affinities are measured using the K_d in nanomolar (nM). The Davis data is particularly focused on kinase inhibitors, making it specific and suitable to be the cross-domain data for testing DTI prediction models. We converted the affinity values into pK_d and also set the threshold as 7 for binarization. We removed the duplicates and the same data pairs in the training data, resulting in 1,374 positives and 13,061 negatives. Our test set has 68 unique drugs and 366 unique targets.

The BIOSNAP dataset is collected from MolTrans (Huang *et al.* 2021b) test sets. We adopted their unseen drug set and unseen protein set for model evaluation. Dataset statistics are provided in Table 1, available as supplementary data at *Bioinformatics* online.

2.2 Model development

We developed a DTI prediction model, which is named Graph Structure-based Drug–Target Interaction prediction method (GS-DTI). Our model takes SMILES of drugs and sequences of targets as input, and the architecture comprises two key components: the drug feature extraction module and the target feature extraction module, which are used to obtain the embeddings of our drugs and targets (Fig. 1a).

2.2.1 Drug feature extraction module

The drug feature extraction module first gets molecular graphs of drug SMILES strings. For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_i\}_{i \in [1, N_v]}$ is the set of nodes (atoms), $\mathcal{E} = \{e_{i,j}\}_{i,j \in [1, N_v]}$ is the set of edges (chemical bonds), and here N_v is the number of nodes. Graphs are then fed into KPGT. KPGT proposed the Line Graph Transformer (LiGhT) structure, which
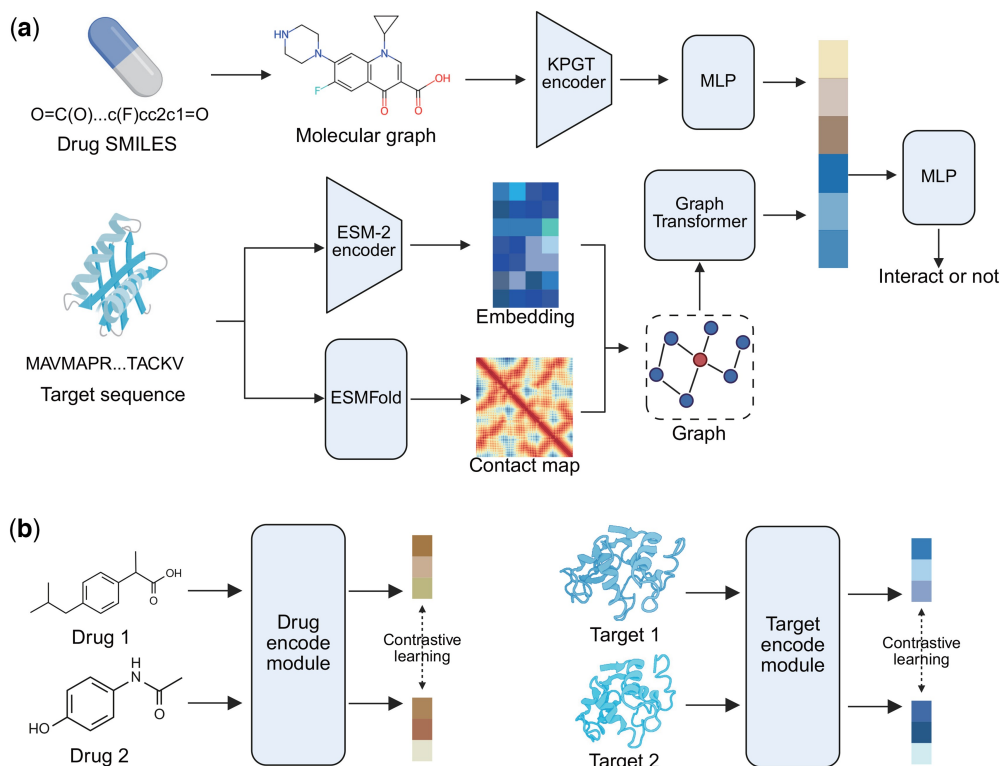
Figure 1. Framework of our proposed method GS-DTI. (a) The architecture of GS-DTI. Both drug and target are represented as graph structures in our model. We also applied pre-trained models to enhance the generalization ability of the model. (b) We designed an intra-contrastive learning between drugs and between targets to help GS-DTI learn generalizable representations.

focuses on chemical bonds and can capture the structural information of molecular graphs. The outputs of KPGT are 2304-dimensional embeddings of the input SMILES strings. The embeddings will be input to a two-layer Multilayer Perceptron (MLP) for information integration, and the finalized embeddings of drugs will be obtained.

2.2.2 Target feature extraction module

We represented the protein as a graph $G = (V, E)$, where V denotes the set of nodes and E represents the edges. To construct the graph, we first predicted the 3D structure of the target protein using ESMFold (Lin *et al.* 2023), an efficient protein structure prediction model without an MSA process. Based on the predicted 3D atomic coordinates, we generated a contact map to define the graph topology. Specifically, each residue in the protein is represented as a node in the graph, and an edge is created between two nodes if the distance between their C_α atoms is less than 8 Å. For the node, we extracted the per-residue embedding of the target from ESM-2 650M (Lin *et al.* 2023) as the feature for each node. The embeddings generated by ESM-2 contain rich evolutionary and functional information about proteins.

After protein graph construction, to extract meaningful representations from the protein graph, we employed a Graph Convolutional Network (GCN) with residual connections, followed by a Graph Multiset Transformer (GMT) (Baek *et al.* 2021) for graph-level embedding aggregation. The GCN captures local structural and relational information by iteratively updating node representations. Specifically, given a protein graph $G = (V, E)$ with initial

node features $\mathbf{H}^{(0)} \in \mathbb{R}^{|V| \times d}$, each GCN layer computes the node features as:

$$\mathbf{H}^{(k)} = \sigma\left(\mathbf{A}\mathbf{H}^{(k-1)}\mathbf{W}^{(k)}\right) + \mathbf{H}^{(k-1)}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{|V| \times |V|}$ is the normalized adjacency matrix, $\mathbf{W}^{(k)} \in \mathbb{R}^{d \times d'}$ is the learnable weight matrix of the k-th layer, $\sigma(\cdot)$ is the ReLU activation function, and the residual connection adds $\mathbf{H}^{(k-1)}$ to the output of the current layer. This residual mechanism improves gradient flow and facilitates the learning of deeper networks. After L layers of GCN, where $L = 3$ in our architecture, the final node embeddings $\mathbf{H}^{(L)}$ are obtained.

To generate a fixed-size graph-level embedding from the variable-sized graph, we then applied a Graph Multiset Transformer (GMT). The GMT aggregates the node embeddings $\mathbf{H}^{(L)}$ into a global representation $\mathbf{z}_G$ while preserving important structural and relational information. Formally, the graph embedding is computed as:

$$\mathbf{z}_G = \text{GMT}\left(\mathbf{H}^{(L)}, \mathbf{A}\right), \tag{2}$$

where $\text{GMT}(\cdot)$ denotes the multiset transformer mechanism, which uses attention to weight and aggregate node features dynamically. The resulting graph embedding $\mathbf{z}_G$ serves as a compact and informative representation of the protein structure. The embeddings will also be input to a 2-layer MLP for information integration to generate the finalized protein embeddings.

At last, the drug embedding and target embedding were combined using a bilinear layer, and the resulting representation was used to predict DTI. The architecture hyperparameters of GS-DTI are listed in Table 2, available as supplementary data at *Bioinformatics* online.

2.2.3 Loss function designed for the DTI prediction task

To address the challenges of data imbalance and the need for discriminative representation learning in DTI prediction, we proposed a novel loss function that integrates Focal loss (Lin *et al.* 2017) for the main DTI task and intra-contrastive learning (Chen *et al.* 2020) (Fig. 1b) as an auxiliary objective.

Given the inherent class imbalance in DTI datasets, we employed the focal loss to focus learning on hard-to-classify samples and mitigate the dominance of easy negatives. The focal loss is defined as:

$$
\mathcal{L}_{\text{focal}} = -\frac{1}{N} \sum_{i=1}^{N} [w(1-\widehat{p}_i)^t y_i \log(\widehat{p}_i) \\
+ (1-w)\widehat{p}_i t(1-y_i) \log(1-\widehat{p}_i)], \quad (3)
$$

where $\widehat{p}_i = \sigma(\widehat{y}_i)$ is the predicted interaction probability from the sigmoid function, y_i is the ground truth label, w balances positive and negative classes, and t adjusts the focus on hard examples.

To further enhance the discriminative power of the learned drug and target representations, we introduce an intra-contrastive learning loss based on a generalized NT-Xent loss. Unlike the original one, which defines positive pairs as two different augmentations of the same instance, our approach leverages domain-specific structural similarity to define positive pairs within a batch. Specifically, for a batch of n samples with normalized embeddings $\mathbf{z}_i \in \mathbb{R}^d$, we first compute the cosine similarity matrix among all embeddings. For each pair (i,j), a positive pair is defined if their structural similarity S_{ij} (e.g., Tanimoto coefficient for drugs or TM-score for proteins) exceeds a given threshold θ. The loss for the batch is then computed as follows:

$$
\sim(\mathbf{z}_i, \mathbf{z}_j) = \frac{\mathbf{z}_i \cdot \mathbf{z}_j}{||\mathbf{z}_i||_2 ||\mathbf{z}_j||_2},
$$
$$
\mathbf{L}_{ij} = \frac{\sim(\mathbf{z}_i, \mathbf{z}_j)}{\tau},
$$
$$
M_{ij} = \begin{cases} 1, & S_{ij} > \theta \text{ and } i \neq j, \\ 0, & \text{otherwise}, \end{cases} \quad (4)
$$

Here, $\mathbf{z}_i$ and $\mathbf{z}_j$ denote the normalized embeddings of samples i and j, respectively. τ is a temperature hyperparameter, S_{ij} is the structural similarity between samples i and j, and θ is the threshold for defining positive pairs.

The generalized NT-Xent contrastive loss is:

$$
\mathcal{L}_{\text{contrast}} = -\frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij}} \sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij} \cdot \log \frac{\exp(\mathbf{L}_{ij})}{\sum_{k=1}^{n} \exp(\mathbf{L}_{ik})}, \quad (5)
$$

where M_{ij} indicates whether the pair (i,j) is considered a positive, and the denominator sums over all samples in the batch for normalization.

The final loss is:

$$
\mathcal{L}_{\text{total}} = \alpha \cdot \mathcal{L}_{\text{focal}} + \beta \cdot \mathcal{L}_{\text{contrast\_drug}} + \gamma \cdot \mathcal{L}_{\text{contrast\_target}}, \quad (6)
$$

where α, β, and γ are hyperparameters controlling the contribution of each component.

2.2.4 Training details

For the training hyperparameters, we applied the Adam optimizer with the initial learning rate of 5×10^{-5} and a weight decay of 1×10^{-4}. The total number of epochs is set to 50, with the early stopping strategy to get the best model during training. A batch size of 64 was used throughout the training. For the focal loss, the parameters were set to $w = 0.5$ and $t = 2$. For the intra-contrastive loss we applied, the structural similarity thresholds (θ) were set to 0.8 for drugs and 0.5 for proteins, with a temperature parameter $\tau = 0.07$. The contributions of the focal loss and contrastive learning components were controlled by the hyperparameters α as 1, β and γ as 0.05.

For hyperparameter selection, we employed a grid search strategy by choosing the best-performing hyperparameter combination on 5-fold cross-validation on the training set. Specifically, we predefined a set of candidate values for each hyperparameter and systematically evaluated all possible combinations (Table 3, available as supplementary data at *Bioinformatics* online). All the experiments were run on four V100 GPU cards with 32GB of memory. For other included models, we used the hyperparameters they provided in the code repositories to train them. As GraphDTA (Nguyen *et al.* 2021) is originally trained for predicting the binding affinity, we changed its loss function to binary cross-entropy to make it a DTI prediction model.

2.3 Similarity estimation

For drug molecules, we applied Tanimoto Similarity to measure the structural similarity between drugs, which can be computed as:

$$
Tanimoto(A, B) = \frac{A * B}{||A||^2 + ||B||^2 - A * B}, \quad (7)
$$

where A and B are the feature vectors of drug A and drug B, respectively, and we adopted the Morgan fingerprint as the vector here.

To measure the structure similarity between proteins, we used the TM-scores calculated by TM-align (Zhang and Skolnick 2005). TM-score is independent of the protein size; it quantifies the similarity between two structures using a score ranging from (0, 1], where a score of 1 indicates a complete match between the structures.

To compute the sequence similarity between proteins, we adopted the Needleman–Wunsch algorithm provided by the "pairwise2" function from the Biopython library. This function optimizes the algorithm so that it can be better parallelized for higher efficiency.

3 Results

To evaluate the validity of our proposed method, we conducted the following experimental analyses. First, we
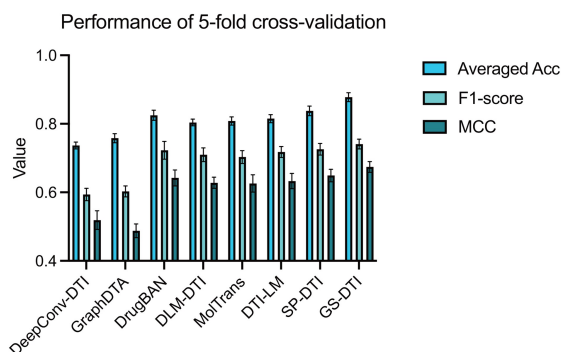
Figure 2. Performance of our model and included baseline methods on 5-fold cross-validation. Averaged Acc is the average value of positive and negative accuracy, indicating the model's overall balanced classification ability across both classes.

performed 5-fold cross-validation to compare our approach with other DTI prediction methods. Next, to assess cross-domain generalizability, we evaluated the model on independent test sets using both overall and cold-start tests. Finally, we applied GS-DTI to identify binding sites for interpretability and to facilitate drug discovery for the BACE1 target.

3.1 GS-DTI generalizes to unseen drugs and targets

We first trained and tested our model on the BindingDB (Liu *et al.* 2007) dataset, which contains 33,238 negative samples and 8,846 positive samples, including 9,833 unique drugs and 1,235 unique proteins. We compared GS-DTI with multiple baseline methods (Lee *et al.* 2019, Nguyen *et al.* 2021, Huang *et al.* 2021b, Bai *et al.* 2023, Lee *et al.* 2024, Ahmed *et al.* 2024, Liu *et al.* 2025) and calculated various evaluation metrics, including Balanced Accuracy, Precision, Recall, F1-score, and MCC. To achieve a fair comparison, all the models were retrained on the same data as ours. On the BindingDB dataset, all the models are tested with the stratified five-fold cross-validation strategy. Considering the negative data is obviously more than the positive data, we applied the stratified cross-validation to ensure each fold maintains the same proportion of positive and negative samples as the original dataset. From Fig. 2, all the models demonstrate stable performance across 5 folds. Our model has the best performance on Averaged Acc, F1-score, and MCC, demonstrating improved discriminative power in imbalanced contexts. Here, Averaged Acc is the averaged value of positive and negative accuracy.

Table 4, available as supplementary data at *Bioinformatics* online reports the average performance of methods on all 5 folds. Here, DeepConv-DTI has very high accuracy on negative data and achieves 0.966, but on positive data, the accuracy is merely 0.512. DeepConv-DTI did not take measures to address the data imbalance, making it not precise in predicting drug–target pairs that are with interactions. All other models except GraphDTA have accuracy on positive data above 0.7, and that of our model is even higher than 0.8, indicating these methods are more effective in handling data with DTI.

As cross-validation is within the in-domain data, all methods achieve decent performance. However, in the real-world application scenario, DTI prediction models are usually applied to drugs and targets that they have not seen in the training data. Therefore, generalization ability is essential for the

model. To demonstrate the advantages of GS-DTI on unseen data, we first trained our model on the full BindingDB dataset, and then tested it on a domain-shift dataset, Davis (Davis *et al.* 2011), with multiple settings. The Davis data is particularly focused on kinase inhibitors, making it appropriate to simulate the model application.

Figure 3a and Table 5, available as supplementary data at *Bioinformatics* online, show the overall performance on the Davis dataset. Our model outperforms all baseline models across all BACC, Precision, Recall, F1-score, and MCC. Since the training data has a data imbalance problem, the negative data is 3 times more than the positive data. Some models are more inclined to learn from negatives. DeepConv-DTI, GraphDTA, and DLM-DTI have good performance on inactive drug–target pairs but are struggling to accurately predict data pairs with DTI, leading to low BACC and MCC. DrugBAN, MolTrans, DTI-LM, SP-DTI, and GS-DTI have a more balanced performance on both types, and our model is more precise and superior to the second-best model, SP-DTI, by about 0.05 on BACC.

Then, we applied three cold start settings, including drug cold start, target cold start, and drug–target pair cold start, on Davis data to comprehensively evaluate the models. For the drug cold start, all the drugs in the Davis dataset have no overlap with the training data. The results are presented in Fig. 3b and Table 6, available as supplementary data at *Bioinformatics* online. This time, GS-DTI still has the best performance across all metrics and achieves a BACC of 0.857. We substantially outperform all other methods by at least 0.08 on MCC. Although DLM-DTI, DTI-LM, and SP-DTI also learned drug information by taking molecule pretrained models' embeddings as input, GS-DTI has contrastive learning to further enhance the discriminative power of the learned drug representations. The cross-domain adaptation strategy enables DrugBAN to predict unseen drugs, making it the second-best model in the drug-cold start test.

In the scenario of the target cold start, we removed all the targets that existed in the training set. We found GS-DTI and SP-DTI outperform other models by a wider margin (Fig. 3c and Table 7, available as supplementary data at *Bioinformatics* online). The reason could be that both GS-DTI and SP-DTI model the structure of proteins and leverage embeddings generated by the protein language model, thus capturing richer structural and functional information than those that just take the pure amino acid sequence or pre-trained embeddings as input. In addition, our model has better BACC, Precision, F1-score, and MCC than SP-DTI, indicating robust performance in handling unseen targets.

For the drug–target pair cold start setting, none of the drugs and targets in this test overlapped with the training set, which allows the models to make inferences on completely unseen data. GS-DTI still wins the first place across all metrics and significantly outperforms other models for at least 0.120 on MCC (Fig. 3d and Table 8, available as supplementary data at *Bioinformatics* online).

We also compared the AUROC (Area Under the Receiver Operating Characteristic Curve) and AUPRC (Area Under the Precision-Recall Curve) of all models under the drug–target pair cold start setting (Fig. 4). GS-DTI has comparable performance with the SOTA methods, demonstrating its ability to distinguish between interacting and non-interacting drug–target pairs. This consistent outperformance across multiple evaluation criteria highlights the robustness and
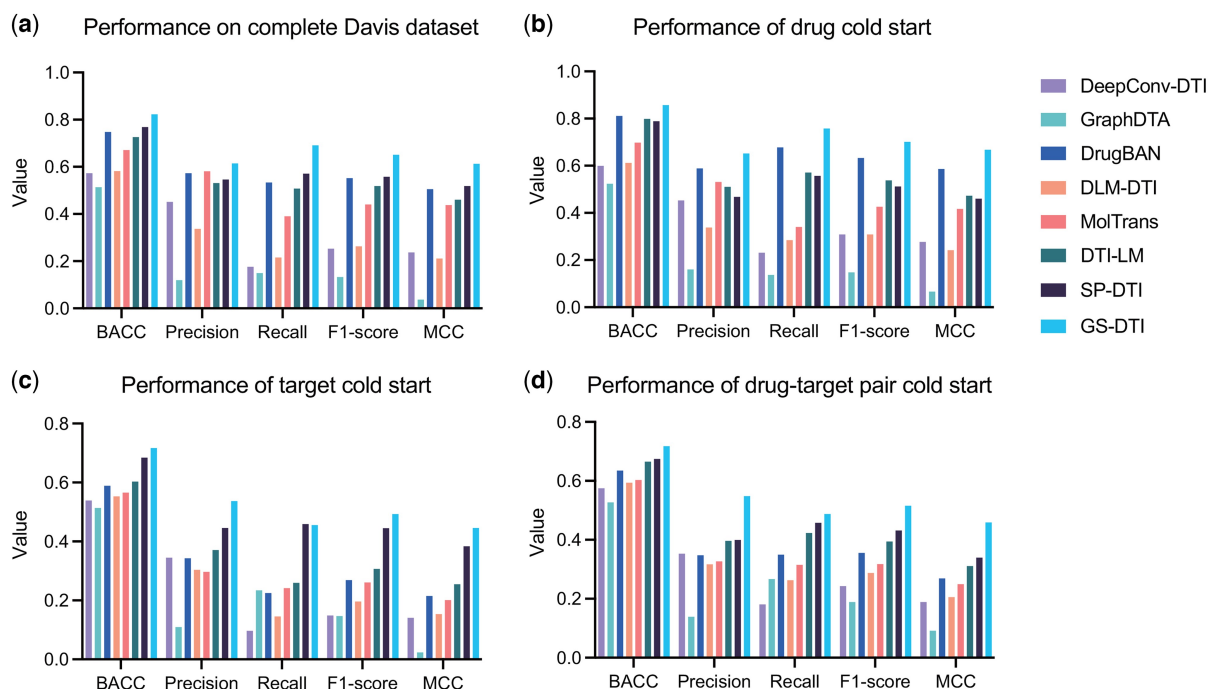
Figure 3. Performance comparison of GS-DTI and baseline methods on the Davis dataset under four evaluation settings: (a) full data, (b) drug cold-start, (c) target cold-start, and (d) drug–target pair cold-start.
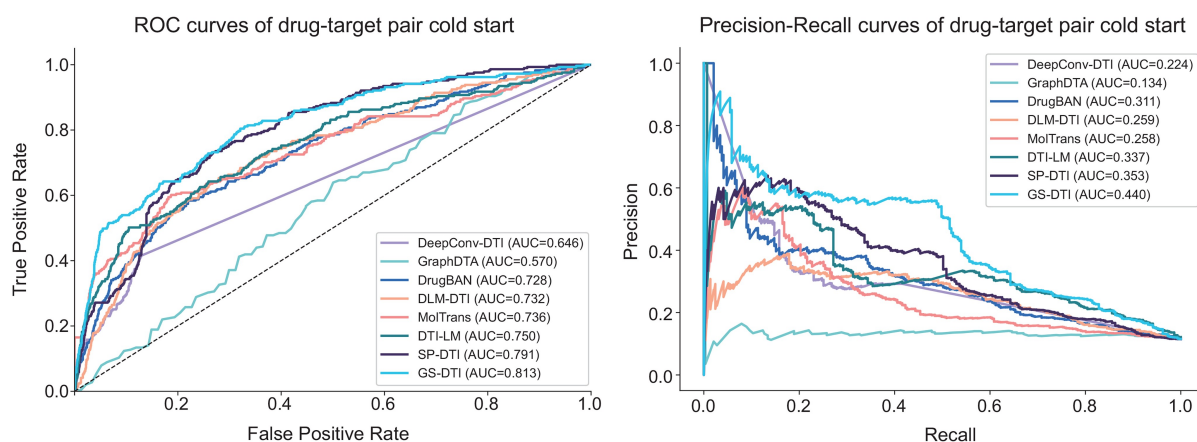

Figure 4. ROC curves and Precision-Recall curves of GS-DTI and baseline methods on the Davis drug–target pair cold-start test.

generalizability of GS-DTI, suggesting that our approach is particularly effective in capturing the underlying patterns necessary for accurate drug–target interaction prediction.

Besides performance comparison, we then evaluated the representations learned by GS-DTI and visualized the concatenated embeddings of drug and target (Fig. 5) using t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton 2008). Before training, the embeddings of active and inactive drug–target pairs are distributed in a scattered and largely overlapping manner. After training, the embeddings become much more structured, with a clear separation between the active and inactive samples. This indicates that the GS-DTI model learned informative representations,

effectively mapping drug–target pairs with different activity labels into distinct regions of the embedding space.

Furthermore, we quantified the epistemic uncertainty of GS-DTI. Specifically, we applied Monte Carlo Dropout (Gal and Ghahramani 2016) and randomly sampled 100 pairs (50 positive and 50 negative) from each of four test settings on the Davis dataset. We performed 50 stochastic forward passes per sample to compute the variance of the model output as the uncertainty measure. This variance quantifies fluctuations in raw prediction scores, with higher values indicating greater model uncertainty. As shown in Fig. 1, available as supplementary data at *Bioinformatics* online, uncertainty was lowest for drug cold-start and modestly higher for target and pair cold-start
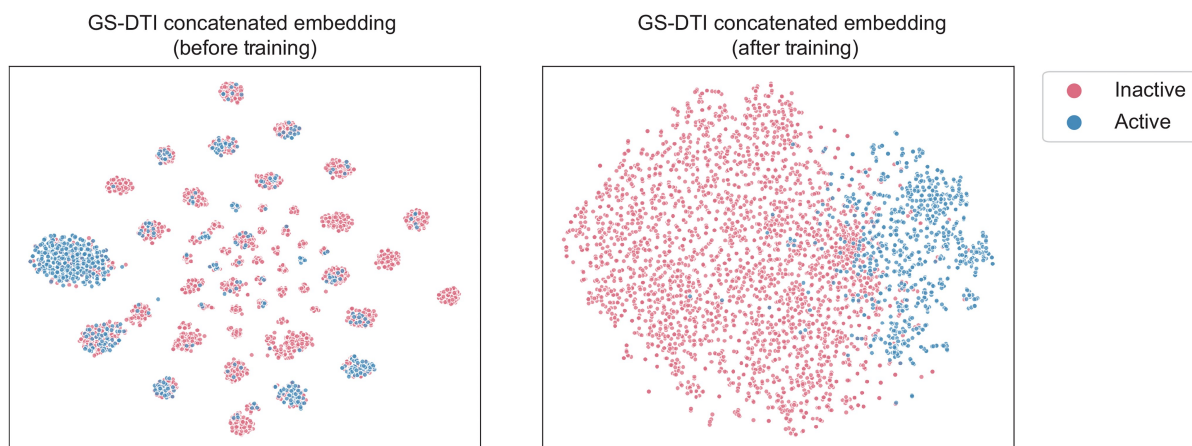
Figure 5. Visualization of concatenated drug–target pair embeddings using t-SNE before and after training. We took the Davis data as input and extracted the embeddings from the output of the bilinear layer.

Table 1. Performance (ROC-AUC) on BIOSNAP unseen drug and protein split tests.[a]

Settings	DeepConv-DTI	GraphDTA	DrugBAN	DLM-DTI	MolTrans	DTI-LM	SP-DTI	GS-DTI
Unseen drugs	0.847 ± 0.009	0.792 ± 0.012	0.872 ± 0.005	0.859 ± 0.009	0.853 ± 0.011	0.874 ± 0.009	0.894 ± 0.009	**0.901 ± 0.010**
Unseen proteins	0.766 ± 0.022	0.650 ± 0.024	0.771 ± 0.024	0.828 ± 0.020	0.770 ± 0.029	0.866 ± 0.019	0.873 ± 0.019	**0.884 ± 0.021**

[a] Bold indicates the best performance.

compared to non-cold-start, with minimal gaps underscoring GS-DTI's robustness in unseen data.

Subsequently, we assessed GS-DTI on the BIOSNAP dataset, which has a balanced data distribution, with drug- and target-cold start settings. Specifically, 20% of the drugs and targets, along with all DTI pairs involving them, were selected as the test set. Table 1 shows that GS-DTI has superior performance to SOTA deep learning baselines under both settings. The results are consistent with those in the Davis dataset, which suggests that GS-DTI can generalize to unseen drugs and proteins. For both drugs and targets, our model takes advantage of large-scale pre-trained models to boost the generalization ability and encode advanced features. Furthermore, considering the core role of structures in the drug–target interaction process, we adopted the structures for both the drug and target to make the model more precise. The combination of these techniques allows our model to handle unseen data well.

3.2 Ablation study to support the effectiveness of modules

In this section, we evaluated the contributions of different modules in GS-DTI by conducting an ablation study on the full Davis dataset. We proposed 4 ablation models. First, to examine the gain brought by incorporating protein structural information, we removed the ESMFold module. In this variant, target features were extracted solely from ESM-2, and an MLP was added to process the embeddings, which were then concatenated with the drug embeddings; other modules remained unchanged. We refer to this model as GS-DTI w/o tertiary structure. Second, we excluded the graph multiset transformer (GMT), utilizing only the GCN for molecular representation (GS-DTI w/o GMT), to assess the enhancement provided by the transformer module. Third, in GS-DTI

w/o KPGT, we replaced the KPGT embeddings with Morgan fingerprints derived from SMILES, allowing us to evaluate the effectiveness of the pre-trained molecular graph encoder. The final ablation model, GS-DTI w/o contrastive learning, was trained without the contrastive learning objective to examine its impact on model performance.

The performance of all ablation models is summarized in Table 2. The results clearly demonstrate that protein structural information is essential for our model. Excluding it leads to a substantial performance drop. This decline is likely because ESM-2 embeddings are averaged across the sequences to obtain the same length features for all sequences, which merely contain sequence-level information. Nevertheless, the graph structure includes the residual-level protein information, implicitly promoting our model to learn the binding sites on targets.

Similarly, GS-DTI w/o GMT shows inferior performance, indicating that removing the graph transformer makes the model too shallow to effectively capture complex structural information.
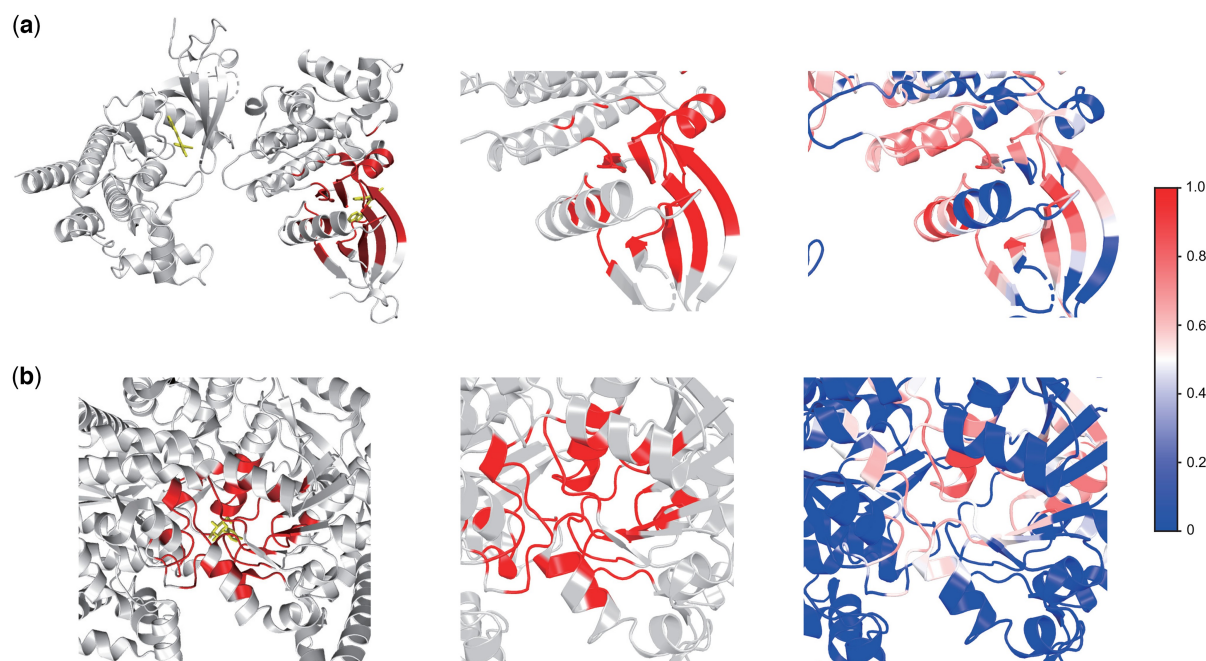
Replacing the KPGT feature with the Morgan fingerprint also results in performance degradation, though not as pronounced. While the Morgan fingerprint does not leverage a pre-trained drug encoder, it still encapsulates the molecule's two-dimensional topological information, which remains useful for this task.

The exclusion of contrastive learning (GS-DTI w/o contrastive learning) leads to a substantial decline in precision, suggesting that it plays a crucial role in enhancing the discriminative power of the learned representations, particularly in distinguishing positive from negative samples.

The above findings offer robust evidence supporting the design choices underlying GS-DTI and highlight the contribution of each module to the model's performance.

Table 2. Ablation study of GS-DTI.[a]

	BACC	Precision	Recall	F1-score	MCC
GS-DTI w/o tertiary structure	0.756 ± 0.016	0.499 ± 0.013	0.598 ± 0.028	0.544 ± 0.022	0.494 ± 0.017
GS-DTI w/o GMT	0.771 ± 0.006	0.558 ± 0.013	0.592 ± 0.024	0.575 ± 0.017	0.529 ± 0.021
GS-DTI w/o KPGT	0.807 ± 0.009	0.538 ± 0.011	0.635 ± 0.018	0.565 ± 0.013	0.522 ± 0.019
GS-DTI w/o contrastive learning	0.792 ± 0.028	0.432 ± 0.046	0.668 ± 0.033	0.539 ± 0.035	0.492 ± 0.045
GS-DTI	$\mathbf{0.823 \pm 0.014}$	$\mathbf{0.615 \pm 0.018}$	$\mathbf{0.691 \pm 0.015}$	$\mathbf{0.651 \pm 0.017}$	$\mathbf{0.613 \pm 0.015}$

[a] Bold indicates the best performance.

Figure 6. Visualization of the tertiary structure of protein–ligand complex and binding pockets. For each subfigure, the left is the protein–ligand complex structure. The middle is the binding pocket structure alone. The right is the binding pocket and surrounding regions. (a) Visualization for a complex with PDB ID 4HGE. (b) Visualization for a complex with PDB ID 3L7B.

3.3 GS-DTI detects binding pockets of targets

To demonstrate that our model is able to detect the binding pocket in the protein, we applied the grad-CAM (Selvaraju *et al.* 2017) on the output of the last graph convolutional layer to identify the important residues that make more contributions to the prediction. The PDBbind database (Wang *et al.* 2005) systematically integrates experimentally measured binding affinity data with three-dimensional structural information of biomolecular complexes from the Protein Data Bank (PDB) (Berman *et al.* 2000). We sampled 2 drug–target pairs in PDBbind (PDB IDs: 4HGE and 3L7B) and checked if the important amino acids are aligned with the pockets of targets. Specifically, we visualized the complex structures, the ground truth pocket, and the GS-DTI predicted important sites.

4HGE (Fig. 6a), a crystal structure of the JAK2 tyrosine kinase (JH1 domain) in complex with the small molecule inhibitor compound 8, providing insights into selective kinase inhibition pertinent to cancer and inflammatory disease therapies (Hanan *et al.* 2012). GS-DTI's high-weight regions closely correspond to the binding pocket regions in the 4HGE crystal structure (Fig. 2, available as supplementary data at *Bioinformatics* online). When the proportion of top-ranked residues selected matches the proportion of annotated binding sites in the protein, the recall—the fraction of true

binding residues correctly identified—is 0.415, indicating the model's ability to reveal the fine-grained drug–target interaction information.

3L7B (Fig. 6b), a crystal structure of glycogen phosphorylase (a key enzyme in glucose metabolism) in complex with the synthetic inhibitor DK3, illustrates the molecular basis for allosteric inhibition of this metabolic enzyme (Tsirkone *et al.* 2010). In the 3L7B structure, DK3 binds to a distinct allosteric site, with its aromatic and hydrophobic moieties fitting tightly into complementary pockets of the enzyme surface, stabilizing an inactive conformation. It could be observed that the residues that make a great contribution to the prediction are highly similar to the binding sites of glycogen phosphorylase (Fig. 3, available as supplementary data at *Bioinformatics* online). Under the same setting, the recall is 0.571, further demonstrating the model's effectiveness in identifying true binding sites.

These alignments suggest that our model effectively captures critical interaction sites, highlighting its potential in implicitly predicting the binding pocket of targets.

3.4 GS-DTI facilitates drug discovery for BACE1

To further test the applicability of GS-DTI, we used it for Beta-site amyloid precursor protein cleaving enzyme 1 (BACE1)

Table 3. Prediction scores of 15 BACE1 inhibitors.

BACE1 inhibitor name	Prediction score
LY2811376	0.9992
Verubecestat	0.9761
LY2886721	0.9990
Epiberberine chloride	1.0000
Tasiamide B-11	0.0046
Lanabecestat	0.9994
Elenbecestat	1.0000
Atabecestat	0.9092
Scoulerine	0.9748
PF-06751979	0.9740
Tasiamide B-9	0.0392
AZD3839 free base	0.6613
LX2343	0.8584
BACE1-IN-5	0.9995
BACE1-IN-4	0.9984

drug discovery by virtual screening. BACE1 is an aspartic-acid protease that plays a critical role in the production of β-amyloid peptides, which accumulate to form amyloid plaques—a hallmark of Alzheimer's disease pathology. The inhibition of BACE1 can reduce β-amyloid generation and has therefore been extensively studied as a potential therapeutic strategy for Alzheimer's disease and related neurodegenerative disorders (Cole and Vassar 2007). We collected the approved drugs in the DrugBank database (Knox *et al.* 2024) and removed the drugs from our training data for the virtual screening of potential BACE1 inhibitors, which resulted in 439 drugs. Notably, BACE1 is an unseen target since it is not in our training data. We also included 15 experimentally verified BACE1 inhibitors to evaluate the recall ability of GS-DTI, and there are a total of 454 molecules for screening.

The predicted results of the 15 inhibitors are listed in Table 3. We found that our model could successfully rediscover 13 out of the 15 (86.7%) inhibitors, and the remaining two are both derivatives of Tasiamide B, a peptide that serves as the Cathepsin D inhibitor, and not specific for binding with BACE1 (Li *et al.* 2019). It's worth noting that 10 of the 15 BACE1 inhibitors were recalled in the top-60 potential drugs by GS-DTI. Additionally, our model predicted that 43/439 drugs in DrugBank have a high probability (confidence score >0.99, Table 9, available as supplementary data at *Bioinformatics* online) to bind to BACE1. We found that among these high-confidence candidates, doxercalciferol (Thiel *et al.* 2023), alfacalcidol (Thiel *et al.* 2023), and penbutolol (Pauls *et al.* 2021) had been verified to be the BACE1 inhibitors. Fosamprenavir is an HIV protease inhibitor, which is a class of drugs that are structurally related to BACE1 inhibitors (Mulato *et al.* 2024). These literatures provide evidence that GS-DTI is effective for BACE1 inhibitor virtual screening.

We also visualized the embeddings of 15 verified inhibitors and 439 drugs for screening (Fig. 4 at *Bioinformatics* online). From the results, 13 out of 15 GS-DTI identified known BACE1 inhibitors cluster closely with the high-confidence candidate drugs. The proximity of their distance indicates that our model effectively captures structural and functional features relevant to BACE1 inhibition. Furthermore, other candidates and predicted non-inhibitors form distinct clusters, suggesting that GS-DTI can robustly separate likely inhibitors from non-inhibitors based on the learned representations. The clustering supports that the embedding space encodes meaningful biochemical information relevant to

BACE1 activity. The above findings demonstrate that GS-DTI can serve as a valuable tool to accelerate the drug development process.

4 Conclusion

In this study, we proposed GS-DTI, a framework that incorporates the structural information of both drugs and proteins for DTI prediction. Comprehensive experimental analysis indicates our proposed method achieved better performance on unseen drugs and targets. Furthermore, by identifying residues that make important contributions to the prediction, our model can provide interpretable insights for the binding pockets. GS-DTI also shows promising potential in discovering novel drug candidates and serves as a powerful tool for drug development.

Our study, like many DTI prediction works, relies on K_d values for labels, which vary due to assay protocols. To mitigate this, we converted K_d to pK_d and binarized the labels, focusing on classification rather than regression. However, some assay-dependent variability may remain and could affect model performance and transferability. In addition, this binary classification may lose granularity by merging weak and non-binders, overlooking affinity nuances for tasks like lead optimization. Future work may consider addressing this by incorporating assay metadata or using more standardized benchmarking datasets. Besides, extending the framework to multi-class frameworks (e.g., categorizing into strong, weak, and non-binders) to preserve finer details from pK_d data is worth investigation.

In the future, we plan to further enhance our model and develop a more comprehensive and automated pipeline for drug discovery. We will also focus on improving the scalability and efficiency of GS-DTI, enabling large-scale virtual screening and accelerating the identification of promising lead compounds.

Author contributions

Qinze Yu (conceptualization [lead], data curation [lead], formal analysis [equal], investigation [lead], methodology [lead], software [equal], validation [supporting], visualization [lead], writing—original draft [lead], writing—review & editing [lead]), Chang Zhou (formal analysis [equal], methodology [supporting], software [equal], validation [lead], writing—original draft [supporting], writing—review & editing [supporting]), Jiyue Jiang (methodology [supporting], resources [supporting], writing—original draft [supporting], writing—review & editing [supporting]), Xiangyu Shi (methodology [supporting], resources [supporting], writing—original draft [supporting], writing—review & editing [supporting]), and Yu Li (conceptualization [supporting], funding acquisition [lead], project administration [lead], resources [lead], supervision [lead], writing—original draft [supporting], writing—review & editing [supporting])

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest: The authors declare that they have no competing interests.

Funding

References

Ahmed KT, Ansari MI, Zhang W. DTI-LM: language model powered drug–target interaction prediction. *Bioinformatics* 2024; **40**:btae533.

Baek J, Kang M, Hwang SJ. Accurate learning of graph representations with graph multiset pooling. In: *Proceedings of the Ninth International Conference on Learning Representations*, May 3-7, Vienna, Austria. ICLR, 2021.

Bai P, Miljković F, John B *et al*. Interpretable bilinear attention network with domain adaptation improves drug–target prediction. *Nat Mach Intell* 2023;**5**:126–36.

Berman HM, Westbrook J, Feng Z *et al*. The protein data bank. *Nucleic Acids Res* 2000;**28**:235–42.

Chen T, Kornblith S, Norouzi M *et al*. A simple framework for contrastive learning of visual representations. In: *Proceedings of the 37th International Conference on Machine Learning*, July 13-18, Vienna, Austria. ICML, 2020.

Chen X, Yan CC, Zhang X *et al*. Drug–target interaction prediction: databases, web servers and computational models. *Brief Bioinform* 2016;**17**:696–712.

Cole SL, Vassar R. The Alzheimer's disease *β*-secretase enzyme, bace1. *Mol Neurodegen* 2007;**2**:1–25.

Davis MI, Hunt JP, Herrgard S *et al*. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* 2011;**29**:1046–51.

Deng J, Yang Z, Ojima I *et al*. Artificial intelligence in drug discovery: applications and techniques. *Brief Bioinform* 2022;**23**:bbab430.

Ding H, Takigawa I, Mamitsuka H *et al*. Similarity-based machine learning methods for predicting drug–target interactions: a brief review. *Brief Bioinform* 2014;**15**:734–47.

Elnaggar A, Heinzinger M, Dallago C *et al*. Prottrans: toward understanding the language of life through self-supervised learning. *IEEE Trans Pattern Anal Mach Intell* 2022;**44**:7112–27.

Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: *Proceedings of the 33rd International Conference on Machine Learning*, June 19-24, New York City, USA. ICML, 2016.

Gu Z, Luo X, Chen J *et al*. Hierarchical graph transformer with contrastive learning for protein function prediction. *Bioinformatics* 2023; **39**:btad410.

Hanan EJ, van Abbema A, Barrett K *et al*. Discovery of potent and selective pyrazolopyrimidine janus kinase 2 inhibitors. *J Med Chem* 2012;**55**:10090–107.

Huang K, Fu T, Gao W *et al*. Therapeutics data commons: machine learning datasets and tasks for drug discovery and development. In: *Proceedings of the Thirty-fifth Conference on Neural Information Processing Systems*, December 6-14. NeurIPS, 2021a.

Huang K, Xiao C, Glass LM *et al*. Moltrans: molecular interaction transformer for drug–target interaction prediction. *Bioinformatics* 2021b;**37**:830–6.

Knox C, Wilson M, Klinger CM *et al*. Drugbank 6.0: the drugbank knowledgebase for 2024. *Nucleic Acids Res* 2024;**52**:D1265–75.

Lee I, Keum J, Nam H. Deepconv-DTI: prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS Comput Biol* 2019;**15**:e1007129.

Lee J, Jun DW, Song I *et al*. DLM-DTI: a dual language model for the prediction of drug-target interaction with hint-based learning. *J Cheminform* 2024;**16**:14.

Li H, Gao Z, Kang L *et al*. TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res* 2006; **34**:W219–24.

Li H, Zhao D, Zeng J. Kpgt: knowledge-guided pre-training of graph transformer for molecular property prediction. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Washington, DC, USA, 2022, 857–67.

Li Z, Bao K, Xu H *et al*. Design, synthesis, and bioactivities of tasiamide b derivatives as cathepsin d inhibitors. *J Pept Sci* 2019;**25**:e3154.

Lin T-Y, Goyal P, Girshick R *et al*. Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017, 2980–8.

Lin Z, Akin H, Rao R *et al*. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023; **379**:1123–30.

Liu S, Liu Y, Xu H *et al*. SP-DTI: subpocket-informed transformer for drug-target interaction prediction. *Bioinformatics* 2025;**41**:btaf011.

Liu T, Lin Y, Wen X *et al*. Bindingdb: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res* 2007;**35**:D198–201.

Mulato A, Lansdon E, Aoyama R *et al*. Preclinical characterization of a non-peptidomimetic HIV protease inhibitor with improved metabolic stability. *Antimicrob Agents Chemother* 2024;**68**:e01373–23.

Nguyen T, Le H, Quinn TP *et al*. Graphdta: predicting drug–target binding affinity with graph neural networks. *Bioinformatics* 2021; **37**:1140–7.

Öztürk H, Özgür A, Ozkirimli E. Deepdta: deep drug–target binding affinity prediction. *Bioinformatics* 2018;**34**:i821–9.

Pauls E, Bayod S, Mateo L *et al*. Identification and drug-induced reversion of molecular signatures of Alzheimer's disease onset and progression in App^{NL-G-F}, App^{NL-F}, and 3xTg-AD mouse models. *Genome Med* 2021;**13**:168.

Pinzi L, Rastelli G. Molecular docking: shifting paradigms in drug discovery. *Int J Mol Sci* 2019;**20**:4331.

Selvaraju RR, Cogswell M, Das A *et al*. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017, 618–26.

Shen J, Yu Q, Chen S *et al*. Unbiased organism-agnostic and highly sensitive signal peptide predictor with deep protein language model. *Nat Comput Sci* 2024;**4**:29–42.

Thiel A, Hermanns C, Lauer AA *et al*. Vitamin D and its analogues: from differences in molecular mechanisms to potential benefits of adapted use in the treatment of Alzheimer's disease. *Nutrients* 2023; **15**:1684.

Tsirkone VG, Tsoukala E, Lamprakis C *et al*. 1-(3-deoxy-3-fluoro-*β*-d-glucopyranosyl) pyrimidine derivatives as inhibitors of glycogen phosphorylase B: kinetic, crystallographic and modelling studies. *Bioorganic & Medicinal Chemistry* 2010;**18**:3413–25.

Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;**9**:2579–605.

Voitsitskyi T, Stratiichuk R, Koleiev I *et al*. 3DProtDTA: a deep learning model for drug-target affinity prediction based on residue-level protein graphs. *RSC Adv* 2023;**13**:10261–72.

Wang R, Fang X, Lu Y *et al*. The pdbbind database: methodologies and updates. *J Med Chem* 2005;**48**:4111–9.

Wang Y, Xia Y, Yan J *et al*. Zerobind: a protein-specific zero-shot predictor with subgraph matching for drug-target interactions. *Nat Commun* 2023;**14**:7861.

Yang H, Chen X, Zuo Y *et al*. MINDg: a drug–target interaction prediction method based on an integrated learning algorithm. *Bioinformatics* 2024;**40**:btae147.

Yu Q, Dong Z, Fan X *et al*. HMD-AMP: protein language-powered hierarchical multi-label deep forest for annotating antimicrobial peptides. arXiv, 2021, preprint: not peer reviewed.

Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005;**33**:2302–9.