MultiPPIMI: a multimodal deep learning framework for predicting interactions between protein-protein interaction targets and modulators

AUTHORS

Heqi Sun[1], Jianmin Wang[4,*], Hongyan Wu[3], Shenggeng Lin[1], Junwei Chen[1], Yi Xiong[1,5,*] and Dong-Qing Wei[1,2,6,*]

[1] State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint International Research Laboratory of Metabolic & Developmental Sciences and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, China
[2] Peng Cheng National Laboratory, Shenzhen, 518055, China
[3] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[4] The Interdisciplinary Graduate Program in Integrative Biotechnology and Translational Medicine, Yonsei University, Incheon 21983, Republic of Korea
[5] Shanghai Artificial Intelligence Laboratory, Shanghai, 200232, China
[6] Zhongjing Research and Industrialization Institute of Chinese Medicine, Nanyang, 473006, China

* To whom correspondence should be addressed. Email: dqwei@sjtu.edu.cn, xiongyi@sjtu.edu.cn, jmwang113@hotmail.com.

Present Address: Dong-Qing Wei, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, P.R. China; Yi Xiong, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, P.R. China.

ABSTRACT

Small molecule modulators of protein-protein interactions (PPIs) have emerged as promising drug candidates with potential applications in anticancer, antiviral, and antimicrobial therapies. While recent virtual screening methods have demonstrated promising results in identifying modulators for specific PPI targets, challenges remain in predicting modulators for novel PPI targets and vice versa. For the first time, we directly predict PPI-modulator interaction through a binary classification task. We construct a benchmark dataset comprising PPIs information and active/inactive modulators data. Moreover, we develop a novel deep learning framework MultiPPIMI that leverages multimodal representations of PPI targets and modulators, and incorporates a bilinear attention network to capture inter-molecule interactions between them. Experiments conducted under both in-domain and cross-domain settings demonstrate the impressive performance and generalization capabilities of MultiPPIMI. The hit rate for identifying modulators of PPIs is greatly improved by combining deep learning with molecular docking. We believe that this work represents a significant step forward in the development of PPI-targeted therapeutics, offering new insights into predicting and understanding the interactions between small molecules and PPIs.

INTRODUCTION

Protein-protein interactions (PPIs) play a critical role in numerous cellular processes, such as cell proliferation, survival and apoptosis (1–4). Dysfunctions in PPIs have been associated with various physio-pathologies including cancer development, infectious diseases and neurological disorders (5). Unlike traditional drug targets with active sites, which might exhibit close resemblance to other regions of the proteome and lead to unintended off-target effects, the broad diversity of PPIs offers a better chance for developing selective and specific drugs (6, 7). Therefore, modulating PPIs with small molecule compounds is regarded as a promising way to treat diseases. However, the low success rate of PPI modulators (inhibitors/stabilizers) in drug development is mainly due to the structural properties of PPIs, which include highly variable interfaces without well-defined binding pockets, and larger interface area (1000-2000 Å2) with high hydrophobicity (6). Due to these properties, PPI modulators are typically large hydrophobic molecules with limited oral bioavailability and cell permeability. Nonetheless, subsequent studies have demonstrated that the hotspot region in PPIs can indeed be targeted by compounds with smaller molecular weight (8–10).

Due to the need for special considerations in identifying interaction hotspots, high-throughput screening for PPI modulators (PPIMs) has a low success rate compared to enzymes or individual protein receptors (6). By contrast, in silico screening of PPIMs can greatly narrow down the search scope of candidate compounds. Structure-based approaches, like molecular docking and molecular dynamics (MD) simulations, are commonly used to screen large compound libraries against PPI partner proteins (11). However, identifying novel scaffolds by docking is still challenging, primarily due to the large and featureless interfaces present in PPI complexes (12).

In recent years, machine learning (ML) has gained popularity in the drug discovery field mainly due to the increased availability of public databases (13, 14). In the field of PPIM discovery, there are several notable databases, including 2P2I-DB (15) which contains over 274 inhibitors that target 27 distinct PPIs, and iPPI-DB (16) which covers over 2378 modulators that interact with 46 PPI families. The emergence of these large datasets has facilitated the development of data-driven methods for screening PPIMs. For example, Hamon et al. explore 40 PPI inhibitors and use DRAGON descriptors with Support Vector Machine (SVM) algorithm to expand the chemical libraries targeting PPIs (17). PPIMpred employs SVM along with 11 molecular features to classify inhibitors for three oncogenic PPIs, namely c-Myc/Max, mdm2/P53 and Bcl2/Bak (18). SMMPPI employs a two-stage classification method based on molecular structure fingerprints (19). In the first stage, it identifies compounds with a high likelihood of inhibiting PPIs. The second stage predicts the inhibitory activity of these compounds against 11 different PPI families. pdCSM-PPI leverages a graph-based compound representation to aid in the discovery of PPIMs and has been applied to 21 distinct PPI targets (20). While numerous studies use ML approaches to screen PPIMs for specific targets, current models have limitations in generalizing to novel PPI targets. Furthermore, these models tend to infer novel modulators by similarity to known ones, resulting in a restricted coverage of the chemical space. Therefore, it is valuable to explore the direct modeling of PPI-modulator interactions (PPIMI) to enhance the generalization capacity of predictive models.

Representation learning of compounds and proteins is crucial for both ML and deep learning (DL) models to perform downstream tasks. Handcrafted features are numerical

representations of molecules that are manually designed to capture structural or physicochemical information. Common features include molecular fingerprints and physicochemical properties. Molecular fingerprints for small molecules are typically binary vectors that encode the presence or absence of substructures or fragments, such as the Morgan fingerprint and PubChem fingerprint (21, 22). In contrast, protein fingerprints are commonly calculated from the amino acid composition or subsequences, such as Amino Acid Composition (AAC) and quasi-sequence fingerprint (23, 24). In addition to molecular fingerprints, physicochemical properties are also important for characterizing molecules, with examples including molecular weight, solubility and lipophilicity for compounds, and aliphaticity, hydrophilicity, and polarity for proteins. More recently, DL techniques can automatically extract features (embeddings) from raw data. These DL approaches have greatly benefited from expressive architectures such as Transformer, which can capture complex and long-range dependencies in samples (25, 26). For instance, Transformer-based protein pre-trained models such as ESM-2, TAPE and ProtTrans have demonstrated outstanding performance in various downstream tasks, including structure and function prediction for proteins (27–29). Similarly for compounds, pre-trained graph neural network (GNN) models (e.g., Uni-Mol, GEM) are capable of learning high-level structural features from atoms and bonds in molecules (30, 31).

The incorporation of multimodal representations for compounds and proteins has led to remarkable improvements in deep learning models on performing downstream tasks, such as compound-protein interaction (CPI) prediction (32–34). As such, the exploration of multimodal data in predicting PPI-modulator interactions holds promise, given the diverse information available. Notably, single modality data can only capture a limited perspective of an entity, potentially leading to indiscriminate results in large datasets. The integration of multiple modalities of compounds and proteins, including their physicochemical properties, sequential and structural information, can generate more comprehensive representations and enhance prediction accuracy.

In the present work, we propose a deep learning framework named MultiPPIMI to predict PPIMI using a multimodal bilinear attention network-based model (Figure 1). The model integrates multimodal representations of compounds and PPI targets to enable explicit

learning of inter-molecule interactions. To achieve this, we extract multimodal representations, which include physicochemical properties and structural embeddings of modulators and PPI targets. These representations are then fed into a pairwise interaction module with a bilinear attention network for capturing local interaction representations, as illustrated in Supplementary Figure S2. Finally, a fully connected layer decodes the local interaction features to predict PPIMI. The main contributions of this study can be summarized as follows:

- We construct a benchmark dataset (including PPI targets information and active/inactive modulators information) of PPIMI from public databases.
- It is the pioneering work to predict PPIMI by formulating it as a binary classification problem, with the focus on screening PPI modulators.
- We experimentally validate the efficacy of multimodal learning and the bilinear attention mechanism in enhancing prediction accuracy and generalization across warm-start and cold-start scenarios.
- We showcase that combining MultiPPIMI with molecular docking can facilitate the efficient identification of lead compounds in large-scale PPI inhibitor screening.

MATERIAL AND METHODS

In PPIMI prediction, the objective is to determine the interaction between a modulator and a PPI target. If the PPI complex involves more than two partners, we randomly select two of them to represent the complex. The amino acid sequence of each partner is denoted as $\mathrm{P_i} = (\mathrm{a_1}, \ldots, \mathrm{a_n})$, where a_j represents one kind of amino acid. Therefore, each PPI target is represented as $(\mathrm{P_1}, \mathrm{P_2})$. Regarding the compound, we use the simplified molecular-input line-entry system (SMILES) representation to calculate physicochemical features, and convert it into a 2D molecular graph for structural embedding calculation. A molecular graph, defined as $G = (V, E)$, consists of a set of vertices (atoms) denoted as V, and a set of edges (chemical bonds) denoted as E. The goal of PPIMI prediction is to train a model M that maps the joint feature representation space to an interaction probability score $p \in [0,1]$.

The framework of MultiPPIMI

The MultiPPIMI framework is designed to directly model PPI-modulator interactions (Figure 1). It takes s PPI- modulators pair as input and generates the probability score of their interaction. This framework consists of four main modules: PPI feature extraction, modulator feature extraction, interaction modelling, and classification. In the PPI feature extraction module, structural embeddings and physicochemical property features are concatenated to generate a PPI representation vector. Similarly, modulator structural embeddings and physicochemical properties are combined to represent each molecule. To model the local interactions between modulators and PPI complexes, a bilinear attention network is used. This network captures the intermolecular interactions and fuses the multimodal representations. Finally, a multilayer perceptron (MLP) is utilized as the classifier to predict the interaction between a candidate modulator and a PPI target.

MultiPPIMI is implemented using Python 3.7.15 and PyTorch 1.8.0 (35). The model is trained using cross-entropy loss with a batch size of 64, and the Adam optimizer with a learning rate of 0.0005. The model is trained for a maximum of 500 epochs. The best-performing model is selected based on the epoch that yielded the highest AUROC score on the validation set. All hyperparameters related to the GraphMVP compound encoder and the bilinear attention module are set to their default values as specified in the original papers. The hyperparameters of batch size, learning rate, and epochs are fine-tuned under the random-split setting and remain consistent across all evaluation scenarios.

Protein-protein interaction representation

For PPI target representation, we combine the structural and physicochemical modality. The structural embeddings are obtained using the pre-trained protein language model ESM2, due to its success in various tasks (36). ESM2 is an unsupervised Transformer model trained on protein sequences from the UniRef database (37). Despite its simple masked language modeling objective during training, ESM2 effectively produces structural features by leveraging the diversity of protein sequences. In this study, we use the 150M ESM2 model as an encoder, generating 640-dimensional embeddings for each protein sequence.

Physicochemical properties are also important as they can influence protein types, structures, and functions. Since protein sequences consist of 20 unique amino acids, each with distinct

chemical and physical properties, we utilize 19 physicochemical features for each amino acid to capture this diversity (Supplementary Table S2). Using the Pfeature tool (38), a 1×19 vector is created for each protein sequence, representing the combined physicochemical properties of all its amino acids.

Compound representation

Compound structural embeddings are computed using the Graph Multi-View Pre-training (GraphMVP) framework (39). This framework employs a graph neural network (GNN)-based multi-view pre-training network to establish the connection between 2D topological views and 3D geometric views of small molecule compounds. The input of GraphMVP is 2D molecular graph, where atoms and bonds are represented as nodes and edges, respectively. Even though the encoder takes only 2D molecular graphs as input, the resulting embeddings effectively incorporate the 3D modality. The output is a 300-dimensional embedding vector for each molecule.

Additionally, physicochemical properties are considered as a complementary modality to the structural view. These features are computed using RDKit (40) (Supplementary Table S1). Consequently, a 1×10 feature vector is generated for each input compound.

Bilinear interaction module

Bilinear attention networks are initially designed for visual question answering (VQA), a task involving multiple modalities (41). These networks can capture interactions between pairs of multimodal input channels, such as image and text, allowing for the integration of diverse modalities and facilitating the learning of effective interaction representations. Bilinear attention networks excel at capturing rich joint information in multimodal data while maintaining computational efficiency comparable to unitary attention mechanisms.

In this study, we use an adapted version of a bilinear attention network from a recent study on CPI prediction (42). This module serves to model non-covalent interactions between modulators and PPI complexes, effectively fusing their multimodal representations. The bilinear attention module consists of two layers (Supplementary Figure S2). The first layer is bilinear interaction map that captures pair-wise attention weights. It takes the

representations of PPI targets and modulators as input, referred to as $\mathbf{H_p} = [\mathbf{h_{1p}}, \mathbf{h_{2p}}, \mathbf{h_{3p}}, \dots \mathbf{h_{Mp}}]$ and $\mathbf{H_c} = [\mathbf{h_{1c}}, \mathbf{h_{2c}}, \mathbf{h_{3c}}, \dots, \mathbf{h_{Nc}}]$, where M and N denote the dimensions of PPI and modulator representations, respectively. Then, a bilinear interaction map is constructed from these representations, resulting in a pairwise interaction matrix $\mathbf{I} \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$ for a single head:

$$\mathbf{I} = \left(\left(\mathbf{1} \cdot \mathbf{q}^{\mathrm{T}}\right) \circ \sigma\left((\mathbf{H}_c)^{\mathrm{T}}\mathbf{U}\right)\right) \cdot \sigma\left(\mathbf{V}^{\mathrm{T}}\mathbf{H}_p\right), \tag{1}$$

which defines the relationship between the learnable weight matrices $\mathbf{U} \in \mathbb{R}^{\mathrm{D_c}\times\mathrm{k}}$ and $\mathbf{V} \in \mathbb{R}^{\mathrm{D_p}\times\mathrm{k}}$, the learnable weight vector $\mathbf{q} \in \mathbb{R}^{k}$, the fixed all-ones vector $\mathbf{1} \in \mathbb{R}^{N}$, and the element-wise product $\circ$. Intuitively, the bilinear interaction maps the representations of PPIs and compounds to a common embedding space using the weight matrices $\mathbf{U}$ and $\mathbf{V}$. The interaction is then learned by applying an element-wise product and the weight vector $\mathbf{q}$.

The second layer is a bilinear pooling layer that extracts a joint representation $\mathbf{f}' \in \mathbb{R}^{k}$ of the PPI-modulator pair from the interaction map $\mathbf{I}$. Specifically, the kth column of $\mathbf{f}'$ is calculated as follows:

$$\begin{aligned}\mathbf{f}'_k &= \sigma(\mathbf{H}_c^T\mathbf{U})_k^T \cdot \mathbf{I} \cdot \sigma\left(\mathbf{H}_p^T\mathbf{V}\right)_k \\ &= \sum_{i=1}^{N}\sum_{j=1}^{M} \mathbf{I}_{i,j}\left(\mathbf{h}_c^i\right)^T\left(\mathbf{U}_k\mathbf{V}_k^T\right)\mathbf{h}_p^j,\end{aligned} \tag{2}$$

where the $\mathbf{U}$ and $\mathbf{V}$ are shared with the interaction map layer. Subsequently, a feature map is generated by applying sum pooling to the joint representation vector:

$$\mathbf{f} = \mathrm{SumPool}(\mathbf{f}', \mathrm{s}), \tag{3}$$

where the $SumPool(\cdot)$ function refers to a non-overlapping 1D sum pooling operation with a stride s. In our approach, the multi-head version of the single pairwise interaction is employed by summing the representation vectors from individual heads. Each head is associated with a specific weight vector $\mathbf{q}$, while the weight matrices $\mathbf{U}$ and $\mathbf{V}$ are shared across all heads.

Classification module

The feature map $\mathbf{f}$ obtained from the bilinear attention module is passed through fully connected layers with the ReLU activation function, as depicted below:

$$\mathrm{output} = fc_3(\mathrm{dropout}(\mathrm{ReLU}(fc_2(\mathrm{dropout}(\mathrm{ReLU}(fc_1(\mathbf{f}))))))) \tag{4}$$

Subsequently, a softmax layer is added after the fully connected layers for the final prediction:

$$\mathrm{softmax}(\mathrm{output_i}) = \frac{\exp(\mathrm{output_i})}{\sum_j \exp(\mathrm{output_j})} \tag{5}$$

The output value represents the probability of interaction between a PPI-modulator pair.

Evaluation strategies and metrics

We evaluate our proposed model and the baseline through a comprehensive analysis using four data split settings. To assess the generalization capability of the model, we employ one warm-start setting and three cold-start scenarios. We use four types of cross validation (CV) as follows:

- S1: Random split setting, where 5-fold CV is performed on PPI-modulator pairs.
- S2: Compound cold-start setting, where 5-fold CV is conducted on compounds. The unique compounds are clustered into five groups using the Butina clustering algorithm, which measures chemical structure similarity with ECFP4 fingerprints and a clustering radius of 0.8. The fold that contains unseen compounds is designated as the test set.
- S3: PPI cold-start setting, where 5-fold CV is conducted on PPI targets.
- S4: Cold pair setting, where 5-fold CV is conducted on both compounds and PPI targets. In this scenario, both the testing compounds and testing targets are unseen in the training set.

The performance of binary classification is evaluated using commonly used metrics, including the area under the receiver-operating characteristic curve (AUROC), the area under the precision-recall curve (AUPR), sensitivity, precision, and specificity. These metrics are calculated as follows:

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

Here, TP, TN, FP and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively.

Different pre-training strategies of GraphMVP

This study compares three self-supervised learning tasks for GraphMVP. The original pre-training objective is defined as follows:

$$L_{GraphMVP} = \alpha_1 \cdot L_C + \alpha_2 \cdot L_G, \quad (6)$$

where α_1, α_2 are weighting coefficients. L_C represents the contrastive loss, which treats 2D-3D pairs of the same molecule as positive and negative otherwise. L_G is the reconstruction loss used for generating 3D conformers from the corresponding 2D topology.

The original pre-training strategy focuses on enhancing the 2D representation by leveraging 3D information (Supplementary Data, section 1.2). To further exploit the 2D topology, we additionally include 2D SSL tasks in the original objective as illustrated in (39). We compare two main categories of 2D SSL tasks: generative and contrastive. This results in two variants of $L_{GraphMVP}$, namely GraphMVP-G and GraphMVP-C, with the following objectives:

$$L_{GraphMVP-G} = L_{GraphMVP} + \alpha_3 \cdot L_{Generative\ 2D-SSL} \quad (7)$$

$$L_{GraphMVP-C} = L_{GraphMVP} + \alpha_3 \cdot L_{Contrastive\ 2D-SSL} \quad (8)$$

RESULTS

Datasets and baseline

Our proposed model is trained and evaluated on the curated DLiP dataset (43), which contains active and inactive small molecule modulators associated with PPI targets from various public databases and published literature (iPPI-DB, 2P2I-DB, TIMBAL, and ChEMBL, and Journal of Medical Chemistry). The protein sequences of the PPI partners are retrieved from the UniProt database (44). With strict data filtering procedures (Supplementary Data, section 1.1), we retain 12,605 distinct compounds and 120 PPI targets. Notably, over 80% of the compounds have a pairwise similarity coefficient less than 0.6 (Supplementary Figure S1) calculated via Tanimoto using ECFP4 fingerprints, indicating low redundancy in our filtered dataset.

To construct a balanced interaction dataset, positive samples are defined as pairs of compounds and PPI targets that interact with each other. To generate negative samples, we adopt a similar approach as a previous study (19). Specifically, for each PPI family, we consider all PPIMs targeting PPIs from the other families as inactive and designate pairs of PPI targets and their corresponding putative inactive PPIMs as negative samples. To avoid assigning inactive modulators incorrectly, we exclude all true active modulators for any given PPI target from the putative inactive pool. To address class imbalance, we perform under-sampling on the negative samples related to each PPI target, resulting in a balanced dataset of 11,630 positive pairs and 11,630 negative pairs. Notably, we exclude experimentally validated inactive PPIMs from the original DLiP dataset, as many PPI targets have insufficient or no corresponding inactive PPIMs.

We compare the performance of MultiPPIMI with a ML-based baseline model. The baseline model employs a Random Forest (RF) classifier based on the concatenated ECFP4 (21) and PseAAC fingerprints (45).

Performance evaluation

We evaluate the performance of MultiPPIMI and the baseline model under four different settings (S1-S4) as shown in Table 1. The evaluations are conducted using 5-fold cross-validation. As expected, both models achieve high prediction performance under setting S1, with comparable AUROC values exceeding 0.994. However, the potential issue of overfitting probably results in such cases of high predictive accuracy.

To address this concern, we conduct further evaluations in three cold-start settings (S2-S4). In all three settings, MultiPPIMI consistently outperforms the baseline model. Notably, S2 exhibits higher average AUROC and AUPR compared to S3, indicating that predicting new modulators is generally easier than predicting new PPI targets. This observation can be attributed to the larger number of compounds available in the training dataset compared to targets. In the most challenging scenario S4, MultiPPIMI achieves a significantly higher AUROC than the RF model (0.6913 versus 0.5788). The performance gap between MultiPPIMI and the baseline is notably larger in cold-start settings than random split setting,

demonstrating the superior generalization capability of MultiPPIMI in scenarios that better represent real-world conditions. This suggests that our model effectively learns more accurate and robust interaction rules.

Ablation study

In this study, we conduct an ablation study to assess the contribution of each main component in MultiPPIMI. We investigate the influence of each feature modality by individually removing them from the input (Figure 2a). The results reveal that the GraphMVP and ESM2 embeddings play a crucial role across all four tasks, showing the largest contribution in terms of AUROC. Moreover, incorporating physicochemical properties positively impacts cold-start predictions (S2-S4) but exhibits minimal effect on the random split setting (S1). This suggests that for simpler tasks involving in-distribution generalization, incorporating overly complex features may be unnecessary.

Additionally, we investigate the effectiveness of bilinear attention by comparing three variants of MultiPPIMI, each differing in the calculation of joint representations between modulators and PPI targets: one-side modulator attention, one-side PPI attention, and linear concatenation (Figure 2a and Figure 2b). The one-side attention approach, inspired by (46), attends solely to the sub-patterns within modulator or PPI representations. We replace the bilinear attention in MultiPPIMI with one-side attention to create the two variants. Another variant involves linear concatenation, which simply combines modulator and PPI representations. The results demonstrate that bilinear attention is the most effective method for capturing interaction information in PPIMI prediction. Interestingly, the differences among the four variants are negligible under S1 but more significant under S2-S4, suggesting that bilinear attention is particularly valuable for enhancing out-of-distribution generalization.

Comparison of pre-training strategies to learn compound embeddings

We investigate the impact of different self-supervised learning (SSL) tasks on compound embeddings by employing various SSL tasks to pre-train GraphMVP on the GEOM dataset (47). Table 2 presents the performance comparison of different pre-training strategies. 'Without pre-train' refers to the GraphMVP encoder without any pre-training, while

'GraphMVP' represents the original pre-training task. Additionally, 'GraphMVP-C', and 'GraphMVP-G' are variants of the original pre-training task.

On average, the pre-trained encoders outperform the approach without pre-training, and the two variant pre-training strategies generally outperform the original approach. The 'GraphMVP' task focuses on mapping 3D geometry to 2D topology (2D-3D SSL), while the two variants include an additional 2D view via the incorporation of 2D SSL. The 2D SSL acts as a regularizer to extract more 2D topological information. By combining 2D-3D SSL and 2D SSL, multiple perspectives of information can be captured. This observation suggests that different modalities complement each other during the pre-training process, similar to previous findings that multimodal pre-training can enhance model generalization compared to single-modal pre-training (33).

Among the three pre-training strategies, 'GraphMVP-C' outperforms the others in most cases, making it the preferred choice in our final model. Importantly, the performance improvement achieved by 'GraphMVP-C' becomes more significant as the task difficulty increases. In the challenging cold-start scenario, the AUROC improvement reaches approximately 2.24%, compared to 0.32% in the random split setting. These findings indicate that for the PPIMI prediction task, pre-training on compound embedding is particularly beneficial for out-of-distribution generalization, and that 2D SSL is complementary to 2D-3D SSL.

Parameter sensitivity analysis

We analyze the impact of three key parameters on the performance of MultiPPIMI: the number of attention heads, the hidden dimension, and the kernel size in the bilinear attention module. The number of attention heads varies from 1 to 10, the hidden dimensions range from 128 to 1024, and the kernel size varies from 2 to 10. Figure 3 illustrates that MultiPPIMI demonstrates robustness to these parameters, with only minor fluctuations observed. Therefore, we retain the default parameters as stated in the original paper (42): 2 attention heads, 512 hidden dimensions, and a kernel size of 3. Moreover, we find that multi-head attention consistently outperforms single head attention. This can be attributed to the fact that multi-head attention models the relationships between modulators and PPIs from

multiple feature spaces, aligning with the diverse non-covalent interaction types observed between atoms and amino acids, such as hydrogen bonds and salt bridges.

MultiPPIMI-based screening of protein-protein interaction inhibitors

Due to the low hit rate of the primary screening of small molecule libraries of PPI targets, MultiPPIMI predictions can be used as a starting point for further virtual screening of PPIs inhibitors (48, 49). The Kelch-like ECH-associated protein 1/nuclear factor erythroid 2-related factor 2 (Keap1/Nrf2) PPI is an important target for the regulation of antioxidant defenses and has been relevant to a wide range of diseases including neurodegenerative, autoimmune and metabolic disorders (49, 50). First, we predict small molecule inhibitors against Keap1/Nrf2 PPI targets by the MultiPPIMI model. Then we evaluate the drug-like space for active inhibitors and compounds with thresholds > 0.7 predicted by MultiPPIMI for the Keap1/Nrf2 PPI by mapping the uniform manifold approximation and projection (UMAP) plots (51) of MACCS fingerprints (52) and Ultrafast Shape Recognition with CREDO Atom Types (USRCAT) fingerprint (53). As shown in UMAP plots (Figure 4a and Figure 4b), the hit molecules share chemical space with the active inhibitor of Keap1/Nrf2 PPI. The hit compounds undergo further virtual screening through molecular docking, using the UCSF DOCK6.9 program (Supplementary Data, section 1.3) (54, 55). The crystal structure of Keap1 protein (PDB ID: 4XMB) is selected for molecular docking (56). PyMOL is applied to create figures (57). As shown in Figure 4c, the hit and reference compounds bind to the binding pocket formed by residues Tyr334, Arg380, Thr338, Arg415, Ile461, Arg483, Gln555, Ser555, and Tyr572. The hit compounds exhibit stronger binding affinity than the reference compound. By combining MultiPPIMI with traditional virtual screening helps to accelerate large-scale PPI inhibitor screening.

Conclusion

In this study, we construct a benchmark dataset of PPI-modulator interactions from a public database. Then we propose MultiPPIMI, a novel deep learning framework that directly model the interactions between PPI targets and modulators. By integrating diverse modality types, including structural information and physicochemical properties, we construct a comprehensive representation that enhances prediction performance. Unlike traditional

virtual screening approaches, our model takes a step further by learning the intermolecular interactions between PPIs and modulators, enabling the discovery of underlying interaction rules.

The systematic evaluations demonstrate the superiority of MultiPPIMI over the baseline model, especially in challenging cold-start scenarios that are commonly encountered in real-world applications. Moreover, our results emphasize the importance of multimodal pre-training of compound embeddings, as it enhances PPIMI prediction, particularly in achieving robust out-of-distribution generalization. Case studies have shown that combining deep learning and traditional virtual screening promises to accelerate large-scale screening of PPI modulators. However, MultiPPIMI relies on high dimensional representations of PPI targets, which may pose a computational challenge as the training dataset expands. In future research, we will explore more effective feature fusion strategies that can reduce dimensionality while preserving crucial information. Moreover, the integration of 3D structural feature information of PPI targets for PPI-regulator interaction prediction along with deep molecular generative models for PPI modulators design (58, 59) will contribute to the discovery of novel PPI modulators.

DATA AVAILABILITY

The datasets and source code are publicly available at https://figshare.com/articles/software/MultiPPIMI/23828124 and https://github.com/sun-heqi/MultiPPIMI.

ACKNOWLEDGEMENTS

We are grateful to Prof. Dong-Qing Wei, Associate Prof. Yi Xiong, Dr. Jianmin Wang, Associate Prof. Hongyan Wu, Shenggeng Lin, Junwei Chen for their valuable suggestions, fruitful discussions, technical support, and feedback.

FUNDING

This work was supported by the National Natural Science Foundation of China [Nos. 32070662 and 61832019].

Conflict of interest statement. None declared.

REFERENCES

1. Wells,J.A. and McClendon,C.L. (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*, **450**, 1001–1009.

2. Fry,D.C. (2012) Small-Molecule Inhibitors of Protein-Protein Interactions: How to Mimic a Protein Partner.

3. Chuderland,D. and Seger,R. (2005) Protein-protein interactions in the regulation of the extracellular signal-regulated kinase. *Mol Biotechnol*, **29**, 57–74.

4. Gao,J., Li,W.X., Feng,S.Q., Yuan,Y.S., Wan,D.F., Han,W. and Yu,Y. (2008) A protein-protein interaction network of transcription factors acting during liver cell proliferation. *Genomics*, **91**.

5. David,A., Razali,R., Wass,M.N. and Sternberg,M.J.E. (2012) Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. *Hum Mutat*, **33**.

6. Ran,X. and Gestwicki,J.E. (2018) Inhibitors of protein–protein interactions (PPIs): an analysis of scaffold choices and buried surface area. *Curr Opin Chem Biol*, **44**.

7. Voter,A.F. and Keck,J.L. (2018) Development of Protein–Protein Interaction Inhibitors for the Treatment of Infectious Diseases. In *Advances in Protein Chemistry and Structural Biology*.Vol. 111.

8. Bogan,A.A. and Thorn,K.S. (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol*, **280**.

9. Clackson,T. and Wells,J.A. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science (1979)*, **267**.

10. Villoutreix,B.O., Labbé,C., Lagorce,D., Laconde,G. and Sperandio,O. (2013) A Leap into the Chemical Space of Protein–Protein Interaction Inhibitors. In *Protein‐Protein Interactions in Drug Discovery*, Methods and Principles in Medicinal Chemistry.pp. 63–83.

11. Sable,R. and Jois,S. (2015) Surfing the protein-protein interaction surface using docking methods: Application to the design of PPI inhibitors. *Molecules*, **20**.

12. Pantsar,T. and Poso,A. (2018) Binding affinity via docking: Fact and fiction. *Molecules*, **23**.
13. Dara,S., Dhamercherla,S., Jadav,S.S., Babu,C.M. and Ahsan,M.J. (2022) Machine Learning in Drug Discovery: A Review. *Artif Intell Rev*, **55**.
14. Vamathevan,J., Clark,D., Czodrowski,P., Dunham,I., Ferran,E., Lee,G., Li,B., Madabhushi,A., Shah,P., Spitzer,M., *et al.* (2019) Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*, **18**, 463–477.
15. Basse,M.-J., Betzi,S., Morelli,X. and Roche,P. (2016) 2P2Idb v2: update of a structural database dedicated to orthosteric modulation of protein–protein interactions. *Database*, **2016**, baw007.
16. Torchet,R., Druart,K., Ruano,L.C., Moine-Franel,A., Borges,H., Doppelt-Azeroual,O., Brancotte,B., Mareuil,F., Nilges,M., Ménager,H., *et al.* (2021) The iPPI-DB initiative: a community-centered database of protein–protein interaction modulators. *Bioinformatics*, **37**, 89–96.
17. Hamon,V., Bourgeas,R., Ducrot,P., Theret,I., Xuereb,L., Basse,M.J., Brunel,J.M., Combes,S., Morelli,X. and Roche,P. (2014) 2P2IHUNTER: a tool for filtering orthosteric protein–protein interaction modulators via a dedicated support vector machine. *J R Soc Interface*, **11**, 20130860.
18. Jana,T., Ghosh,A., Das Mandal,S., Banerjee,R. and Saha,S. (2017) PPIMpred: a web server for high-throughput screening of small molecules targeting protein–protein interaction. *R Soc Open Sci*, **4**, 160501.
19. Gupta,P. and Mohanty,D. (2021) SMMPPI: A machine learning-based approach for prediction of modulators of protein-protein interactions and its application for identification of novel inhibitors for RBD:hACE2 interactions in SARS-CoV-2. *Brief Bioinform*, **22**.
20. Rodrigues,C.H.M., Pires,D.E.V. and Ascher,D.B. (2021) PdCSM-PPI: Using Graph-Based Signatures to Identify Protein-Protein Interaction Inhibitors. *J Chem Inf Model*, **61**, 5438–5445.
21. Rogers,D. and Hahn,M. (2010) Extended-connectivity fingerprints. *J Chem Inf Model*, **50**.

22. Kim,S., Chen,J., Cheng,T., Gindulyte,A., He,J., He,S., Li,Q., Shoemaker,B.A., Thiessen,P.A., Yu,B., *et al.* (2019) PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res*, **47**.
23. Reczko,M. and Bohr,H. (1994) The DEF data base of sequence based protein fold class predictions. *Nucleic Acids Res*, **22**.
24. Chou,K.C. (2000) Prediction of protein subcellular locations by incorporating quasi-sequence-order effect. *Biochem Biophys Res Commun*, **278**.
25. Vaswani,A., Shazeer,N., Parmar,N., Uszkoreit,J., Jones,L., Gomez,A.N., Kaiser,L. and Polosukhin,I. (2017) Attention Is All You Need.
26. Brown,T.B., Mann,B., Ryder,N., Subbiah,M., Kaplan,J., Dhariwal,P., Neelakantan,A., Shyam,P., Sastry,G., Askell,A., *et al.* (2020) Language models are few-shot learners. In *Advances in Neural Information Processing Systems*.Vol. 2020- December.
27. Lin,Z., Akin,H., Rao,R., Hie,B., Zhu,Z., Lu,W., Smetanin,N., Verkuil,R., Kabeli,O., Shmueli,Y., *et al.* (2023) Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science (1979)*, **379**.
28. Rao,R., Bhattacharya,N., Thomas,N., Duan,Y., Chen,X., Canny,J., Abbeel,P. and Song,Y.S. (2019) Evaluating Protein Transfer Learning with TAPE.
29. Elnaggar,A., Heinzinger,M., Dallago,C., Rehawi,G., Wang,Y., Jones,L., Gibbs,T., Feher,T., Angerer,C., Steinegger,M., *et al.* (2022) ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Trans Pattern Anal Mach Intell*, **44**.
30. Zhou,G., Gao,Z., Ding,Q., Zheng,H., Xu,H., Wei,Z., Zhang,L. and Ke,G. Uni-Mol: A Universal 3D Molecular Representation Learning Framework.
31. Fang,X., Liu,L., Lei,J., He,D., Zhang,S., Zhou,J., Wang,F., Wu,H. and Wang,H. (2022) Geometry-enhanced molecular representation learning for property prediction. *Nat Mach Intell*, **4**, 127–134.
32. Yang,X., Niu,Z., Liu,Y., Song,B., Luins,W., Zeng,L. and Zeng,X. (2022) Modality-DTA: Multimodality Fusion Strategy for drug-Target Affinity Prediction. *IEEE/ACM Trans Comput Biol Bioinform*, 10.1109/TCBB.2022.3205282.
33. You,Y. and Shen,Y. (2022) Cross-modality and self-supervised protein embedding for compound-protein affinity and contact prediction. *Bioinformatics*, **38**, ii68–ii74.
34. Zhou,D., Xu,Z., Li,W., Xie,X. and Peng,S. (2021) MultiDTI: drug–target interaction prediction based on multi-modal representation learning to bridge the gap between

new chemical entities and known heterogeneous network. *Bioinformatics*, **37**, 4485–4492.
35. Paszke,A., Gross,S., Massa,F., Lerer,A., Bradbury,J., Chanan,G., Killeen,T., Lin,Z., Gimelshein,N., Antiga,L., *et al.* (2019) PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*.Vol. 32.
36. Lin,Z., Akin,H., Rao,R., Hie,B., Zhu,Z., Lu,W., Smetanin,N., Verkuil,R., Kabeli,O., Shmueli,Y., *et al.* (2021) Evolutionary-scale prediction of atomic level protein structure with a language model. 10.1101/2022.07.20.500902.
37. Suzek,B.E., Wang,Y., Huang,H., McGarvey,P.B. and Wu,C.H. (2015) UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, **31**.
38. Pande,A., Patiyal,S., Lathwal,A., Arora,C., Kaur,D., Dhall,A., Mishra,G., Kaur,H., Sharma,N., Jain,S., *et al.* (2023) Pfeature: A Tool for Computing Wide Range of Protein Features and Building Prediction Models. *Journal of Computational Biology*, **30**.
39. Liu,S., Wang,H., Liu,W., Lasenby,J., Guo,H. and Tang,J. PRE-TRAINING MOLECULAR GRAPH REPRESENTATION WITH 3D GEOMETRY.
40. Landrum,G. (2021) RDKit: Open-Source Cheminformatics Software. *Http://Www.Rdkit.Org/*.
41. Kim,J.H., Jun,J. and Zhang,B.T. (2018) Bilinear attention networks. In *Advances in Neural Information Processing Systems*.Vol. 2018- December.
42. Bai,P., Miljković,F., John,B. and Lu,H. (2023) Interpretable bilinear attention network with domain adaptation improves drug–target prediction. *Nat Mach Intell*, 10.1038/s42256-022-00605-1.
43. Ikeda,K., Maezawa,Y., Yonezawa,T., Shimizu,Y., Tashiro,T., Kanai,S., Sugaya,N., Masuda,Y., Inoue,N., Niimi,T., *et al.* (2023) DLiP-PPI library: An integrated chemical database of small-to-medium-sized molecules targeting protein–protein interactions. *Front Chem*, **10**.
44. Bateman,A., Martin,M.J., Orchard,S., Magrane,M., Agivetova,R., Ahmad,S., Alpi,E., Bowler-Barnett,E.H., Britto,R. and Bursteinas,B. (2021) UniProt: the universal protein knowledgebase in 2021 %J Nucleic Acids Research. *Nucleic Acids Res*, **49**.
45. Cao,D.S., Xu,Q.S. and Liang,Y.Z. (2013) Propy: A tool to generate various modes of Chou's PseAAC. *Bioinformatics*, **29**.

46. Tsubaki,M., Tomii,K. and Sese,J. (2019) Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*, **35**.
47. Axelrod,S. and Gómez-Bombarelli,R. (2022) GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Sci Data*, **9**.
48. Shimizu,Y., Yonezawa,T., Bao,Y., Sakamoto,J., Yokogawa,M., Furuya,T., Osawa,M. and Ikeda,K. (2023) Applying deep learning to iterative screening of medium-sized molecules for protein–protein interaction-targeted drug discovery. *Chemical Communications*, 10.1039/D3CC01283B.
49. Guo,W., Wisniewski,J.A. and Ji,H. (2014) Hot spot-based design of small-molecule inhibitors for protein-protein interactions. *Bioorg Med Chem Lett*, **24**.
50. Lee,S. and Hu,L. (2020) Nrf2 activation through the inhibition of Keap1–Nrf2 protein–protein interaction. *Medicinal Chemistry Research*, **29**.
51. McInnes,L., Healy,J., Saul,N. and Großberger,L. (2018) UMAP: Uniform Manifold Approximation and Projection. *J Open Source Softw*, **3**.
52. Durant,J.L., Leland,B.A., Henry,D.R. and Nourse,J.G. (2002) Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci*, **42**.
53. Schreyer,A.M. and Blundell,T. (2012) USRCAT: Real-time ultrafast shape recognition with pharmacophoric constraints. *J Cheminform*, **4**.
54. Allen,W.J., Fochtman,B.C., Balius,T.E. and Rizzo,R.C. (2017) Customizable de novo design strategies for DOCK: Application to HIVgp41 and other therapeutic targets. *J Comput Chem*, **38**.
55. Allen,W.J., Balius,T.E., Mukherjee,S., Brozell,S.R., Moustakas,D.T., Lang,P.T., Case,D.A., Kuntz,I.D. and Rizzo,R.C. (2015) DOCK 6: Impact of new features and current docking performance. *J Comput Chem*, **36**.
56. Jain,A.D., Potteti,H., Richardson,B.G., Kingsley,L., Luciano,J.P., Ryuzoji,A.F., Lee,H., Krunic,A., Mesecar,A.D., Reddy,S.P., *et al.* (2015) Probing the structural requirements of non-electrophilic naphthalene-based Nrf2 activators. *Eur J Med Chem*, **103**.
57. Delano,W.L. (2002) The PyMOL Molecular Graphics System. *CCP4 Newsletter on protein crystallography*, **40**.

58. Wang,J., Chu,Y., Mao,J., Jeon,H.N., Jin,H., Zeb,A., Jang,Y., Cho,K.H., Song,T. and No,K.T. (2022) De novo molecular design with deep molecular generative models for PPI inhibitors. *Brief Bioinform*, **23**.

59. Ohue,M., Kojima,Y. and Kosugi,T. (2023) molecules Communication Generating Potential Protein-Protein Interaction Inhibitor Molecules Based on Physicochemical Properties. 10.3390/molecules28155652.

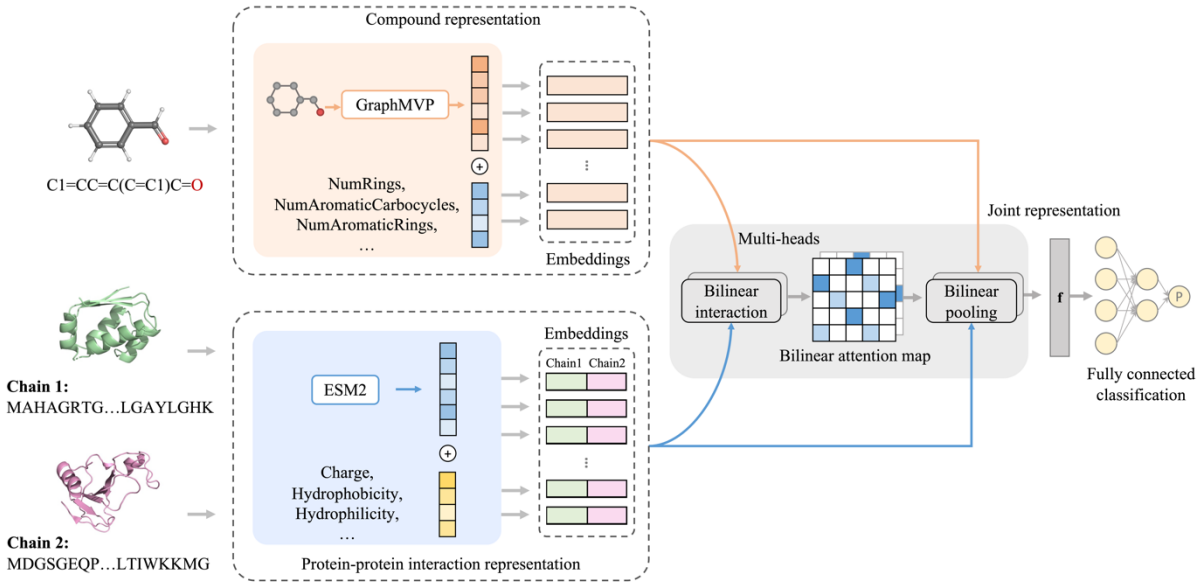

Figure 1. Overview of the MultiPPIMI framework. The modulator and PPI targets are encoded by using compound and protein encoders, respectively. The compound representation consists of the concatenation of pre-trained graph embeddings and physicochemical properties. The PPI representation is constructed by concatenating ESM2 embeddings and physicochemical features of the two chains. The modulator and PPI representations are then inputted into a bilinear attention network, enabling the learning of local interactions. The joint representation **f** is decoded using a fully connected network to predict the probability *p* of PPIMI.

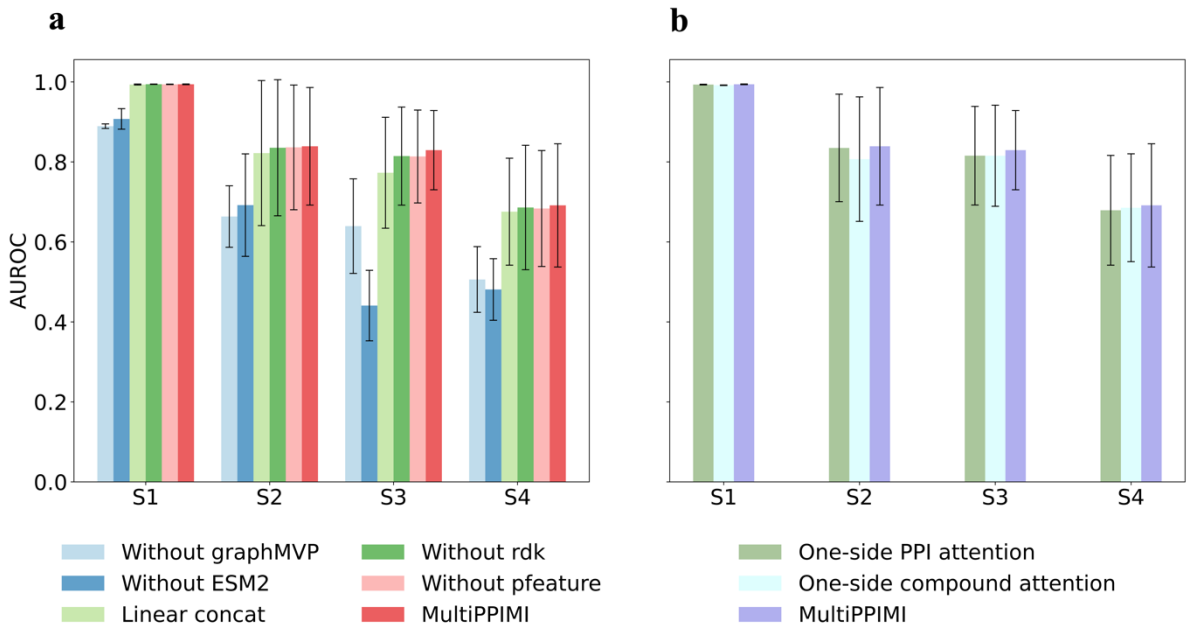

Figure 2. Ablation study with 5-fold cross validation across different data split settings. a. Ablation study on the key components of MultiPPIMI. b. Ablation study on bilinear attention module.

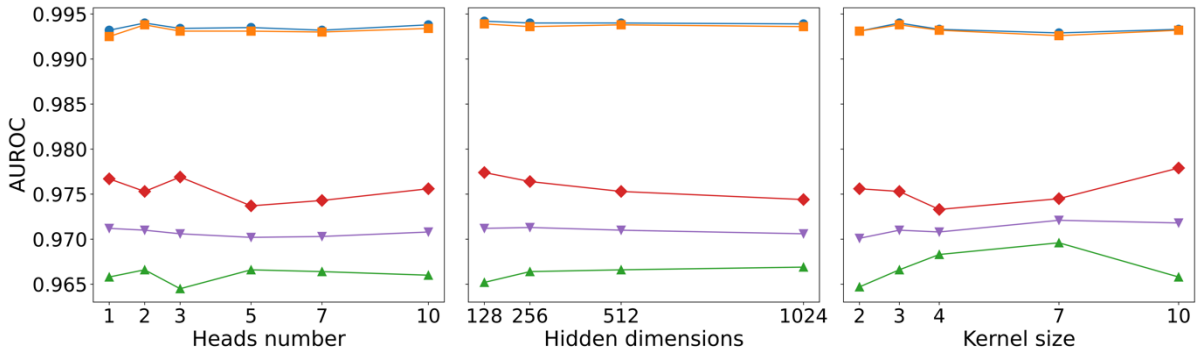

Figure 3. Parameter sensitivity analysis with 5-fold cross validation under random split.

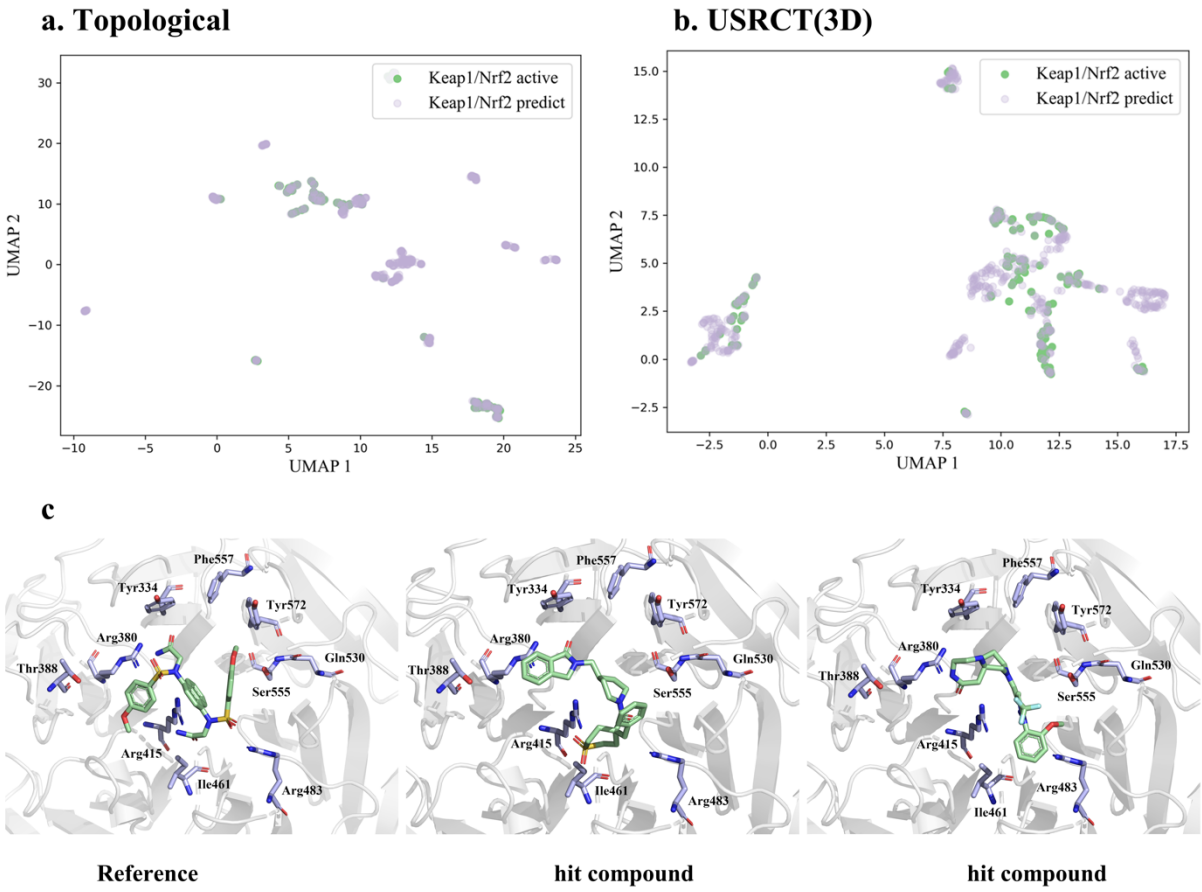

Figure 4. The UMAP visualization of the distribution of Morgan (a) fingerprint and USRCAT (b) fingerprint in two-dimension space (violet: predict compounds; green: active inhibitors). (c) Binding pose of the Keap1 protein with reference compound (PDB ID: 4XMB) and binding poses of the hit compounds using docking.

Table 1. Performance comparison with 5-fold cross validation across various data split settings. The best results for each task are marked **bold**.

Setting	Model	AUROC	AUPR	Sensitivity	Precision	Specificity
S1	MultiPPIMI	0.9940±0.0009	0.9938±0.0009	0.9753±0.0048	0.9666±0.0057	0.9663±0.0062
	RF	0.9953±0.0008	0.9954±0.0009	0.9726±0.0101	0.9679±0.0092	0.9676±0.0099
S2	MultiPPIMI	0.8391±0.1470	0.8366±0.1569	0.8816±0.0417	0.7443±0.1236	0.6553±0.2835
	RF	0.7753±0.1148	0.7945±0.1317	0.8355±0.0702	0.6813±0.1058	0.5631±0.2383
S3	MultiPPIMI	0.8294±0.0991	0.8130±0.1156	0.9336±0.0266	0.6917±0.0913	0.5628±0.1975
	RF	0.6534±0.1267	0.7034±0.0919	0.8969±0.1181	0.5764±0.0718	0.2946±0.2566
S4	MultiPPIMI	0.6913±0.1540	0.6856±0.1435	0.9290±0.0573	0.6177±0.0877	0.3845±0.2600
	RF	0.5788±0.0642	0.5959±0.0855	0.9812±0.0206	0.5109±0.0167	0.0547±0.0631

Table 2. Performance comparison of MultiPPIMI with different pre-training strategies. The best performance in each task is indicated in **bold**.

	S1	**S2**	**S3**	**S4**
Without pre-train	0.9908±0.0009	0.8328±0.1227	0.8222±0.1139	0.6689±0.1541
GraphMVP	0.9938±0.0015	0.8424±0.1411	0.8114±0.1299	0.6574±0.1460
GraphMVP-C	0.9940±0.0009	0.8391±0.1470	0.8294±0.0991	0.6913±0.1540
GraphMVP-G	0.9922±0.0009	0.8476±0.1338	0.8284±0.1045	0.6707±0.1459