A Multimodal Deep Learning Framework for Predicting PPI-Modulator Interactions

Heqi Sun[1], Jianmin Wang[2,*], Hongyan Wu[3], Shenggeng Lin[1], Junwei Chen[1], Jinghua Wei[4], Shuai Lv[1], Yi Xiong[1,5,*], Dong-Qing Wei[1,6,7,*]

[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] *The Interdisciplinary Graduate Program in Integrative Biotechnology and Translational Medicine, Yonsei University, Incheon 21983, Republic of Korea*

[3] *Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China*

[4] *Department of Chemistry, University of Toronto, Toronto M5R 0A3, Canada*

[5] *Shanghai Artificial Intelligence Laboratory, Shanghai, 200232, China*

[6] *Peng Cheng National Laboratory, Shenzhen, 518055, China*

[7] *Zhongjing Research and Industrialization Institute of Chinese Medicine, Nanyang, 473006, China*

ABSTRACT

Protein-protein interactions (PPIs) are essential for many biological processes and diseases. Small molecule modulators of PPIs have emerged as promising drug candidates. However, existing computational methods for identifying PPI modulators rely on similarity to known ones, while neglecting PPI target information and the principles of molecular interaction. Consequently, these methods cannot handle PPI targets with few or no known modulators. To address this challenge, we construct a benchmark dataset that contains PPI targets, modulators, and their interaction relationship. We propose MultiPPIMI, a deep learning framework for predicting the interactions between any PPI-modulator pair. Our model integrates multimodal representations of PPI targets and modulators, and uses a bilinear attention network to capture inter-molecular interactions. We demonstrate that MultiPPIMI achieves an average AUROC of 0.7866 under three cold-start scenarios, and 0.994 under warm-start scenario. Furthermore, we show that combining MultiPPIMI with molecular simulations improves the hit rate for screening inhibitors of Keap1/Nrf2 PPI interactions. This work provides a new way to discover PPI modulators for existing and emerging diseases.

Keywords: PPI-modulator interaction, PPI modulator discovery, deep learning, multimodal representations

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 may share similarities with other regions of the proteome and cause unwanted off-target effects, the diversity of PPIs offers a better chance for developing selective and specific drugs [6,7]. Therefore, modulating PPIs with small molecules is a promising strategy to treat diseases. However, identifying PPI modulators (inhibitors/stabilizers) is challenging due to the special structural properties of PPIs. For example, they have highly variable interfaces without well-defined binding pockets, and larger interface area (1000-2000 Å2) with high hydrophobicity [6]. Because of these properties, PPI modulators are often large hydrophobic molecules with poor oral bioavailability and cell permeability. Despite this, many studies have shown that hotspot regions in PPIs, which contribute most to the binding affinity, can be targeted by compounds with lower molecular weight [8–10].

High-throughput screening is a conventional approach for identifying PPI modulators. Nevertheless, it has a low success rate compared to enzymes or individual protein receptors, due to the need for special considerations to identify interaction hotspots [6]. In contrast, *in silico* screening of PPI modulators (PPIMs) can greatly narrow the search space for candidate modulators. Structure-based approaches, such as molecular docking and molecular dynamics simulations, are commonly used to screen large compound libraries against PPI partner proteins [11]. However, finding novel scaffolds by docking is still challenging, mainly due to the large and featureless interfaces of PPI complexes [12].

Machine learning (ML) or deep learning (DL) has become a powerful tool for drug discovery, owing to the availability of large public databases [13,14]. Among them, several databases focus on PPIMs, such as 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. These datasets enable the application of data-driven methods for screening potential PPIMs. For example, Hamon *et al.* use DRAGON descriptors and Support Vector Machine (SVM) algorithm to identify novel PPI inhibitors based on 40 known ones [17]. PPIMpred applies SVM with 11 molecular features to predict inhibitors for three oncogenic PPIs[18]. SMMPPI adopts a two-stage classification method to predict PPIMs based on molecular structure fingerprints [19]. pdCSM-PPI utilizes a graph-based compound representation to facilitate the discovery of PPIMs for 21 different PPI targets [20]. However, existing methods predict modulators based on similarity to known ones, while ignoring PPI target information and the interaction between PPI targets and modulators. Consequently. these approaches are not suitable for novel or under-studied PPI targets that lack sufficient known modulators. Therefore, it is necessary to develop generalizable approaches that can directly predict the interactions between any PPI target-modulator pairs.

Effective representations of compounds and proteins are essential for predicting PPI target-modulator interactions (PPIMI). DL approaches can effectively learn the representations (embeddings) from raw sequential or structural data of compounds and proteins, using expressive neural network architectures such as Transformer. 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 [21–23]. Similarly for compounds, pre-trained graph neural network (GNN) models (e.g., Uni-Mol, GEM) are capable of learning high-level structural representations from atoms and bonds in molecules [24,25]. On the other hand, 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. However, structural representations alone may not be sufficient to capture the complex interactions between PPI targets and modulators. Integrating multimodal representations for proteins and compounds can generate more comprehensive representations and enhance prediction accuracy. This has been demonstrated in DL models for compound-protein

interaction (CPI) prediction [26–28]. As such, the exploration of multimodal data in predicting PPIMI holds promise.

In this study, we propose a novel deep learning framework named MultiPPIMI to predict PPI target-modulator interactions (**Figure 1**). The model integrates multimodal representations of modulators and PPI targets, including structural embeddings and physicochemical properties, to comprehensively capture their characteristics. The model also uses bilinear attention networks to explicitly learn the inter-molecule interactions between PPI targets and modulators. The main contributions of this study are as follows:

- We construct a benchmark dataset that contains information on PPI targets, modulators, and their interaction relationship.
- We are the first to predict PPI target-modulator interactions by formulating it as a binary classification problem, with the focus on screening PPI modulators.
- We experimentally validate the efficacy of multimodal representations and the bilinear attention mechanism in enhancing prediction accuracy and generalization capability across warm-start and cold-start scenarios.
- We show that combining MultiPPIMI with molecular simulations improves the hit rate for screening inhibitors of Keap1/Nrf2 PPI interactions.

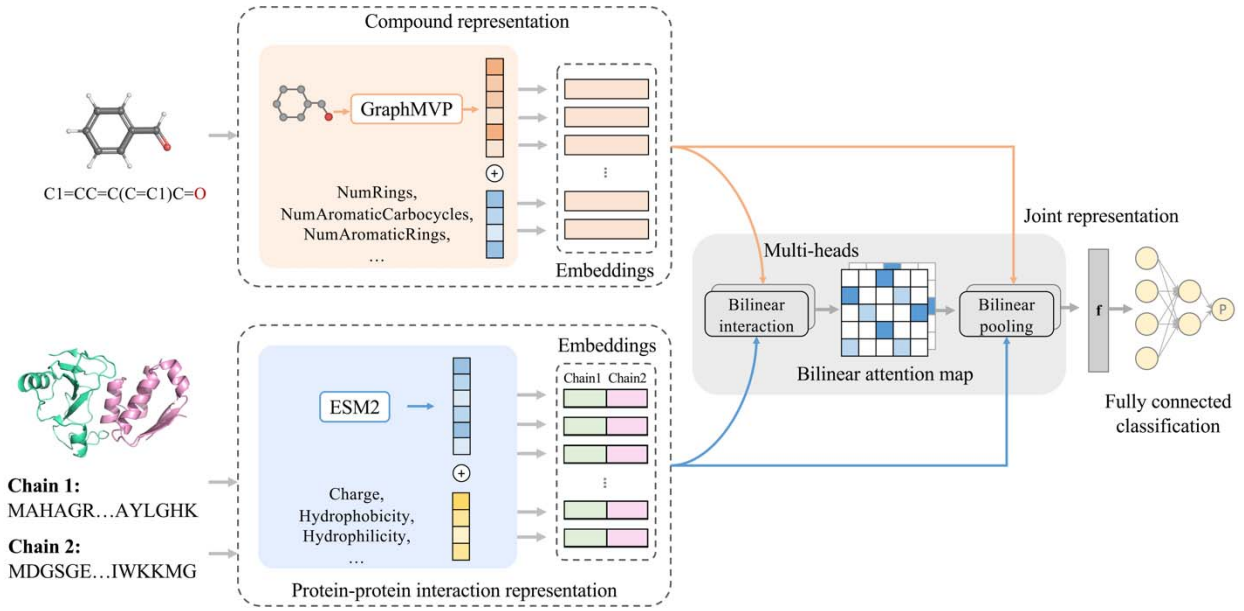

Figure 1. Overview of the MultiPPIMI framework. The modulator representation consists of the concatenation of pre-trained structural embeddings and physicochemical properties. The PPI representation is constructed by concatenating ESM2 embeddings and physicochemical features of two chains. The modulator and PPI representations are input into a bilinear attention network, enabling the learning of inter-molecular interactions. The joint representation **f** is decoded using a fully connected network to predict the probability *p* of PPI target-modulator interactions.

RESULTS AND DISCUSSION

Datasets and baselines. To evaluate our proposed model for PPIMI prediction, we construct a benchmark dataset based on a public database DLiP [29], which contains active and inactive small modulators of PPI targets. The protein sequences of PPI partners are retrieved from the UniProt database [30]. With strict data filtering procedures (Supporting Information, section 1), we retain 12,605 distinct modulators and 120 PPI targets. Notably, over 80% of the modulators have a pairwise similarity coefficient less than 0.6 (Figure S1) calculated via Tanimoto using ECFP4 fingerprints, indicating a low redundancy in our filtered dataset.

To construct a balanced PPIMI dataset, positive samples are defined as pairs of modulators

and PPI targets that interact with each other. To generate negative samples, we adopt a similar approach as a previous study [19]. For each PPI family, we assume all modulators that interact with the other PPI families are inactive. To avoid mis-assigning inactive modulators, we exclude all putative negative modulators that overlap with true active modulators. Negative samples are defined as pairs of PPI targets and putative inactive modulators. To address class imbalance, we perform under-sampling on the negative samples, 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 to reduce noise, because many PPI targets do not have validated 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 [31] for modulator, and PseAAC fingerprints [32] for PPI partners.

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 model is notably larger in cold-start settings than warm-start setting, demonstrating the superior generalization capability of MultiPPIMI in scenarios that better represent real-world conditions. This suggests that our model more effectively learns the interaction rules.

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
	RandomForest	**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**
	RandomForest	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**
	RandomForest	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**
	RandomForest	0.5788±0.0642	0.5959±0.0855	**0.9812±0.0206**	0.5109±0.0167	0.0547±0.0631

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 warm-start 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 [33], 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 uses 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.

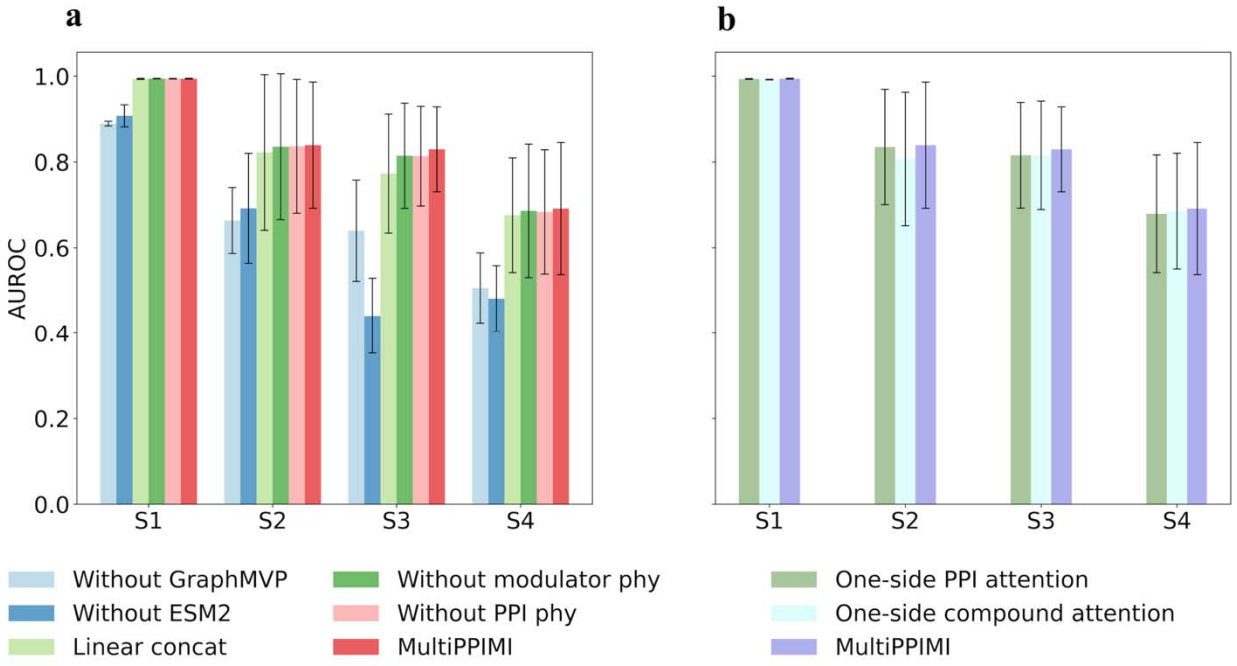

Figure 2. Ablation study with 5-Fold cross-validation across different data split settings. (a) Ablation study on the key components of MultiPPIMI. 'phy' denotes physicochemical features. (b) Ablation study on bilinear attention module.

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 [34]. **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 two variants of the original pre-training task (See MATERIALS AND METHODS).

On average, the pre-trained encoders outperform the approach without pre-training, and the

two variant pre-training strategies 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 stage, similar to previous findings that multimodal pre-training can enhance model generalization compared to single-modal pre-training [27].

Among the three pre-training strategies, 'GraphMVP-C' outperforms the other variants 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.

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

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 are robust to these parameters, with only minor fluctuations observed. Therefore, we retain the default parameters as stated in the original paper [35]: 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 can model the interaction between modulators and PPI targets from multiple feature spaces, which is consistent with the diverse non-covalent interaction types observed between atoms and amino acids, such as hydrogen bonds and salt bridges.

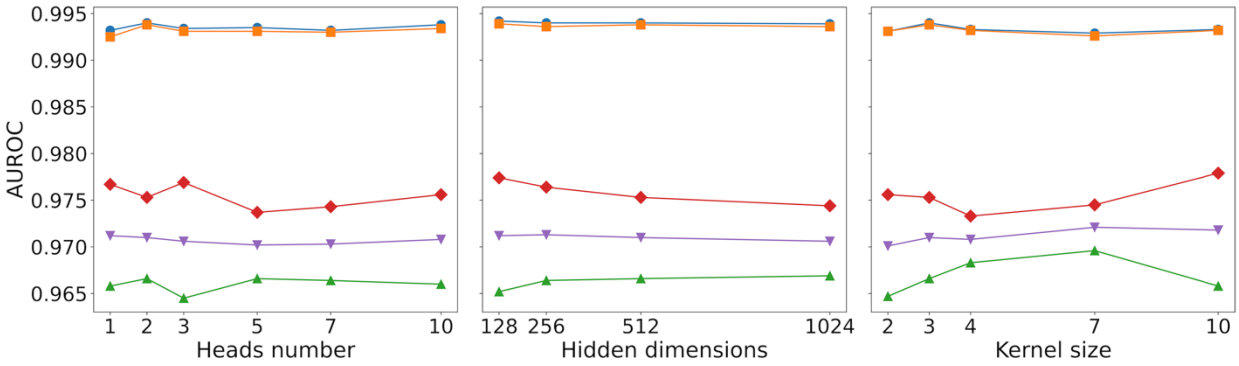

Figure 3. Parameter sensitivity analysis. Key parameters of MultiPPIMI are tested with 5-fold cross-validation under random split setting.

MultiPPIMI-based screening for protein-protein interaction inhibitors. Due to the low hit rate of the primary screening of small molecular libraries against PPI targets, MultiPPIMI predictions can be used as a starting point for further virtual screening PPIs inhibitors [36,37]. 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 [37,38]. First, we predict inhibitors against Keap1/Nrf2 interactions by using MultiPPIMI. Then we evaluate the drug-like space of the active inhibitors of Keap1/Nrf2 and the MultiPPIMI-predicted

inhibitors (probability > 0.7), by mapping the uniform manifold approximation and projection (UMAP) plots [39] of MACCS fingerprints [40] and Ultrafast Shape Recognition with CREDO Atom Types (USRCAT) fingerprint [41]. As shown in UMAP plots (**Figure 4a** and **Figure 4b**), the hit molecules share chemical space with the active inhibitors of Keap1/Nrf2 PPI target. The hit compounds undergo further virtual screening through molecular docking, using the UCSF DOCK6.9 program (Supporting Information, section 3) [42,43]. The crystal structure of Keap1 protein (PDB ID: 4XMB) is selected for molecular docking [44]. PyMOL is applied to create figures [45]. 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.

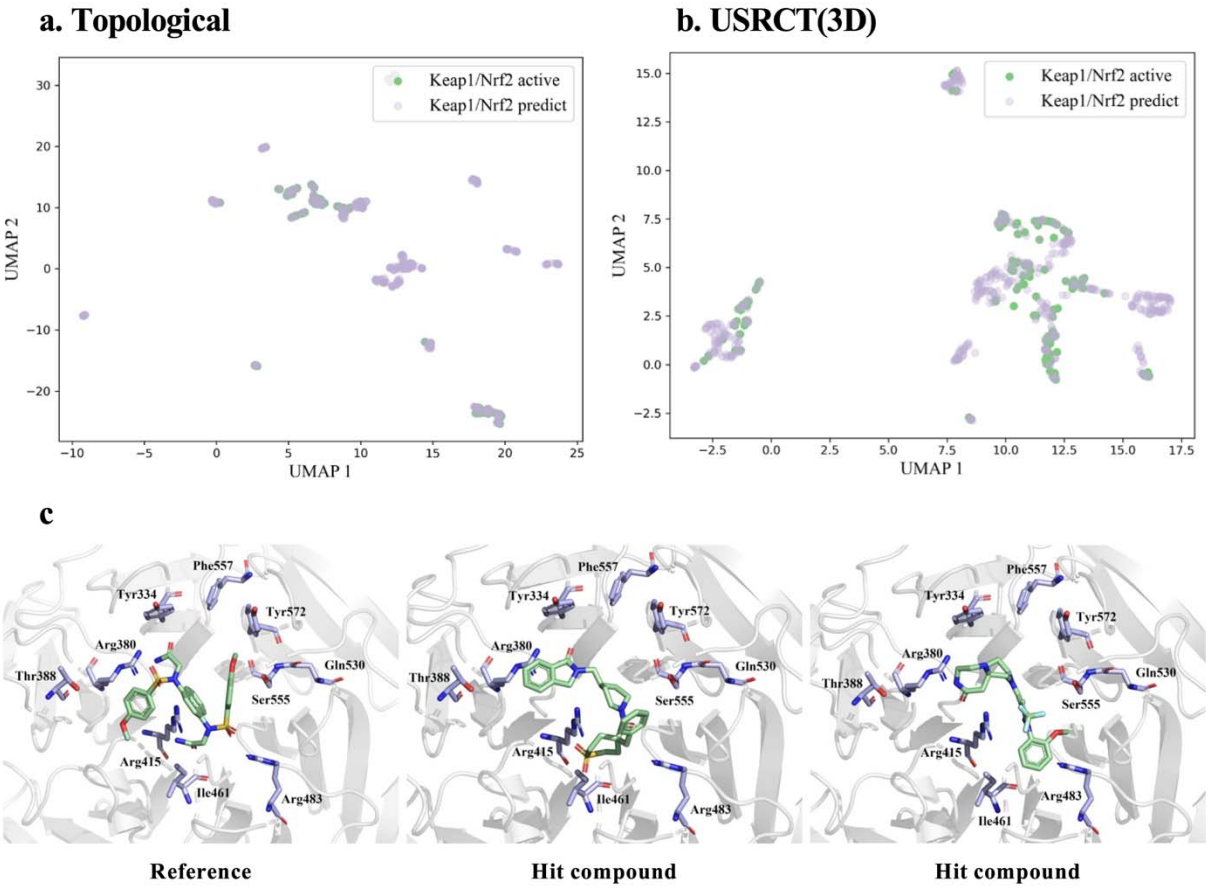

Figure 4. The UMAP visualization of the distribution of (a) MACCS fingerprint and the (b) USRCAT fingerprint in two-dimension space (violet: predicted inhibitors; green: active inhibitors). (c) Binding pose of the Keap1 protein with the reference compound (PDB ID: 4XMB) and binding poses of the hit compounds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

CONCLUSION

In this study, we construct a benchmark dataset of PPI-modulator interactions from public databases and propose MultiPPIMI, a novel deep learning framework for predicting the interactions between PPI target-modulator pairs. Our model integrates multimodal representations of modulators and PPI targets, including structural embeddings and physicochemical properties, to capture the diversity of both entities. Moreover, our model employs a bilinear attention network to learn the local interactions between PPI targets and modulators, which enables the extraction of the generalized interaction rules.

We conduct systematic evaluations to demonstrate the superiority of MultiPPIMI over the baseline model, especially in challenging cold-start scenarios that are frequently encountered in real-world applications. Furthermore, our results highlight the importance of multimodal pre-training of compound embeddings in enhancing PPIMI prediction, particularly in achieving robust out-of-distribution generalization. Case studies have shown that combining MultiPPIMI with molecular simulations promises to improve hit rate of 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 essential information. Additionally, exploiting 3D structural information of PPI targets and integrating MultiPPIMI with deep molecular generative models [46,47] may facilitate the discovery of novel PPI modulators.

MATERIALS AND METHODS

Problem formulation. 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 modulator, 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 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 [48]. 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 [49]. ESM2 is an unsupervised Transformer model trained on protein sequences from the UniRef database [50]. 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 [51], 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 [52]. This framework employs a GNN-based multi-view pre-training network to establish the connection between 2D topological views and 3D geometric views of small 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 [53] (Table S1). A 1×10 feature vector is generated for each input compound.

Bilinear interaction module. Bilinear attention networks are initially designed for visual question answering, a task involving multiple modalities [54]. 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 [35]. This module serves to model non-covalent interactions between modulators and PPI targets. The bilinear attention module consists of two layers. 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:

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

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:

$$\mathbf{f}'_k = \sigma(\mathbf{H}_c^T\mathbf{U})_k^T \cdot \mathbf{I} \cdot \sigma\left(\mathbf{H}_p^T\mathbf{V}\right)_k$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbf{I}_{i,j} \left(\mathbf{h}_c^i\right)^T (\mathbf{U}_k \mathbf{V}_k^T) \mathbf{h}_p^j, \quad (2)$$

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:

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

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:

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

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

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

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

Evaluation strategies and metrics. We evaluate MultiPPIMI and the baseline model using four data split settings that reflect different levels of generalization capability. We use one warm-start setting and three cold-start scenarios, resulting in our types of cross validation (CV):

- S1: Random split setting, where 5-fold CV is performed on PPI-modulator pairs.

- S2: Modulator cold-start setting, where 5-fold CV is conducted on modulators. The unique compounds are clustered into five groups based on their chemical structure similarity using the Butina clustering algorithm with ECFP4 fingerprints and a clustering radius of 0.8.
- 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 area under the receiver-operating characteristic curve (AUROC), area under the precision-recall curve (AUPR), sensitivity, precision, and specificity. These metrics are calculated as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$
$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

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

$$L_{GraphMVP} = \alpha_1 \cdot L_C + \alpha_2 \cdot L_G, \tag{6}$$

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 (Supporting Information, section 2). To further exploit the 2D topology, we additionally include 2D SSL tasks in the original objective as illustrated in [52]. 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:

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

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

ASSOCIATED CONTENT

Data Availability Statement

All the source code, software, and datasets in this publicationis freely available under an MIT license at https://github.com/sun-heqi/MultiPPIMI. The benchmark dataset is stored in the directory "data", the four data split scenarios are stored in "data/folds".

Supporting information

The following file is available free of charge.

Supporting Information: (1) Pre-processing of the DLiP curation dataset, (2) GraphMVP pre-training task, (3) Docking protocol; Figure S1 illustrates similarity of modulator pairs; Table S1 and Table S2 summarizes the physiochemical properties used by MultiPPIMI for small molecules and proteins, respectively (PDF)

AUTHOR INFORMATION

Corresponding Authors

* Dong-Qing Wei – Email: dqwei@sjtu.edu.cn

* Yi Xiong – Email: xiongyi@sjtu.edu.cn

* Jianmin Wang – Email: jmwang113@hotmail.com

Author Contributions

Heqi Sun: conceptualization, formal analysis, writing original draft. Jianmin Wang: conceptualization, data analysis, writing review and editing. Hongyan Wu: supervision. Shenggeng Lin: writing review and editing. Junwei Chen: writing review and editing. Jinghua Wei: data analysis. Shuai Lv: data analysis. Yi Xiong: supervision, resources, writing review and editing. Dong-Qing Wei: supervision, resources, writing review and supervision.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

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

REFERENCES

(1) Wells, J. A.; McClendon, C. L. Reaching for High-Hanging Fruit in Drug Discovery at Protein-Protein Interfaces. *Nature*. Nature Publishing Group December 13, 2007, pp 1001–1009. https://doi.org/10.1038/nature06526.

(2) Fry, D. C. *Small-Molecule Inhibitors of Protein-Protein Interactions: How to Mimic a Protein Partner*; 2012; Vol. 18.

(3) Chuderland, D.; Seger, R. Protein-Protein Interactions in the Regulation of the Extracellular Signal-Regulated Kinase. *Mol Biotechnol* **2005**, *29* (1), 57–74. https://doi.org/10.1385/MB:29:1:57.

(4) Gao, J.; Li, W. X.; Feng, S. Q.; Yuan, Y. S.; Wan, D. F.; Han, W.; Yu, Y. A Protein-Protein Interaction Network of Transcription Factors Acting during Liver Cell Proliferation. *Genomics* **2008**, *91* (4). https://doi.org/10.1016/j.ygeno.2007.12.007.

(5) David, A.; Razali, R.; Wass, M. N.; Sternberg, M. J. E. Protein-Protein Interaction Sites Are Hot Spots for Disease-Associated Nonsynonymous SNPs. *Hum Mutat* **2012**, *33* (2). https://doi.org/10.1002/humu.21656.

(6) Ran, X.; Gestwicki, J. E. Inhibitors of Protein–Protein Interactions (PPIs): An Analysis of Scaffold Choices and Buried Surface Area. *Current Opinion in Chemical Biology*. 2018. https://doi.org/10.1016/j.cbpa.2018.06.004.

(7) Voter, A. F.; Keck, J. L. Development of Protein–Protein Interaction Inhibitors for the Treatment of Infectious Diseases. In *Advances in Protein Chemistry and Structural Biology*; 2018; Vol. 111. https://doi.org/10.1016/bs.apcsb.2017.07.005.

(8) Bogan, A. A.; Thorn, K. S. Anatomy of Hot Spots in Protein Interfaces. *J Mol Biol* **1998**, *280* (1). https://doi.org/10.1006/jmbi.1998.1843.

(9) Clackson, T.; Wells, J. A. A Hot Spot of Binding Energy in a Hormone-Receptor Interface. *Science (1979)* **1995**, *267* (5196). https://doi.org/10.1126/science.7529940.

(10) Villoutreix, B. O.; Labbé, C.; Lagorce, D.; Laconde, G.; Sperandio, O. A Leap into the Chemical Space of Protein–Protein Interaction Inhibitors. In *Protein‐Protein Interactions in Drug Discovery*; Methods and Principles in Medicinal Chemistry; 2013; pp 63–83. https://doi.org/https://doi.org/10.1002/9783527648207.ch4.

(11) Sable, R.; Jois, S. Surfing the Protein-Protein Interaction Surface Using Docking Methods: Application to the Design of PPI Inhibitors. *Molecules*. 2015. https://doi.org/10.3390/molecules200611569.

(12) Pantsar, T.; Poso, A. Binding Affinity via Docking: Fact and Fiction. *Molecules*. 2018. https://doi.org/10.3390/molecules23081899.

(13) Dara, S.; Dhamercherla, S.; Jadav, S. S.; Babu, C. M.; Ahsan, M. J. Machine Learning in Drug Discovery: A Review. *Artif Intell Rev* **2022**, *55* (3). https://doi.org/10.1007/s10462-021-10058-4.

(14) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of Machine Learning in Drug Discovery and Development. *Nature Reviews Drug Discovery*. Nature Publishing Group June 1, 2019, pp 463–477. https://doi.org/10.1038/s41573-019-0024-5.

(15) Basse, M.-J.; Betzi, S.; Morelli, X.; Roche, P. 2P2Idb v2: Update of a Structural Database Dedicated to Orthosteric Modulation of Protein–Protein Interactions. *Database* **2016**, *2016*, baw007. https://doi.org/10.1093/database/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.; Sperandio, O. The IPPI-DB Initiative: A Community-Centered Database of Protein–Protein Interaction Modulators. *Bioinformatics* **2021**, *37* (1), 89–96. https://doi.org/10.1093/bioinformatics/btaa1091.

(17) Hamon, V.; Bourgeas, R.; Ducrot, P.; Theret, I.; Xuereb, L.; Basse, M. J.; Brunel, J. M.; Combes, S.; Morelli, X.; Roche, P. 2P2IHUNTER: A Tool for Filtering Orthosteric Protein–Protein Interaction Modulators via a Dedicated Support Vector Machine. *J R Soc Interface* **2014**, *11* (90), 20130860. https://doi.org/10.1098/rsif.2013.0860.

(18) Jana, T.; Ghosh, A.; Das Mandal, S.; Banerjee, R.; Saha, S. PPIMpred: A Web Server for High-Throughput Screening of Small Molecules Targeting Protein–Protein Interaction. *R Soc Open Sci* **2017**, *4* (4), 160501. https://doi.org/10.1098/rsos.160501.

(19) Gupta, P.; Mohanty, D. 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* **2021**, *22* (5). https://doi.org/10.1093/bib/bbab111.

(20) Rodrigues, C. H. M.; Pires, D. E. V.; Ascher, D. B. PdCSM-PPI: Using Graph-Based Signatures to Identify Protein-Protein Interaction Inhibitors. *J Chem Inf Model* **2021**, *61* (11), 5438–5445. https://doi.org/10.1021/acs.jcim.1c01135.

(21) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science (1979)* **2023**, *379* (6637). https://doi.org/10.1126/science.ade2574.

(22) Rao, R.; Bhattacharya, N.; Thomas, N.; Duan, Y.; Chen, X.; Canny, J.; Abbeel, P.; Song, Y. S. Evaluating Protein Transfer Learning with TAPE. **2019**.

(23) Elnaggar, A.; Heinzinger, M.; Dallago, C.; Rehawi, G.; Wang, Y.; Jones, L.; Gibbs, T.; Feher, T.; Angerer, C.; Steinegger, M.; Bhowmik, D.; Rost, B. ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Trans Pattern Anal Mach Intell* **2022**, *44* (10). https://doi.org/10.1109/TPAMI.2021.3095381.

(24) Zhou, G.; Gao, Z.; Ding, Q.; Zheng, H.; Xu, H.; Wei, Z.; Zhang, L.; Ke, G. *Uni-Mol: A Universal 3D Molecular Representation Learning Framework*. https://github.com/dptech-corp/Uni-Mol.

(25) Fang, X.; Liu, L.; Lei, J.; He, D.; Zhang, S.; Zhou, J.; Wang, F.; Wu, H.; Wang, H. Geometry-Enhanced Molecular Representation Learning for Property Prediction. *Nat Mach Intell* **2022**, *4* (2), 127–134. https://doi.org/10.1038/s42256-021-00438-4.

(26) Yang, X.; Niu, Z.; Liu, Y.; Song, B.; Luins, W.; Zeng, L.; Zeng, X. Modality-DTA: Multimodality Fusion Strategy for Drug-Target Affinity Prediction. *IEEE/ACM Trans Comput Biol Bioinform* **2022**. https://doi.org/10.1109/TCBB.2022.3205282.

(27) You, Y.; Shen, Y. Cross-Modality and Self-Supervised Protein Embedding for Compound-Protein Affinity and Contact Prediction. *Bioinformatics* **2022**, *38* (2), ii68–ii74. https://doi.org/10.1093/bioinformatics/btac470.

(28) Zhou, D.; Xu, Z.; Li, W.; Xie, X.; Peng, S. MultiDTI: Drug–Target Interaction Prediction Based on Multi-Modal Representation Learning to Bridge the Gap between New Chemical Entities and Known Heterogeneous Network. *Bioinformatics* **2021**, *37* (23), 4485–4492. https://doi.org/10.1093/bioinformatics/btab473.

(29) Ikeda, K.; Maezawa, Y.; Yonezawa, T.; Shimizu, Y.; Tashiro, T.; Kanai, S.; Sugaya, N.; Masuda, Y.; Inoue, N.; Niimi, T.; Masuya, K.; Mizuguchi, K.; Furuya, T.; Osawa, M. DLiP-PPI Library: An Integrated Chemical Database of Small-to-Medium-Sized Molecules Targeting Protein–Protein Interactions. *Front Chem* **2023**, *10*. https://doi.org/10.3389/fchem.2022.1090643.

(30) Bateman, A.; Martin, M. J.; Orchard, S.; Magrane, M.; Agivetova, R.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E. H.; Britto, R.; Bursteinas, B. UniProt: The Universal Protein Knowledgebase in 2021 %J Nucleic Acids Research. *Nucleic Acids Res* **2021**, *49* (D1).

(31) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J Chem Inf Model* **2010**, *50* (5). https://doi.org/10.1021/ci100050t.

(32) Cao, D. S.; Xu, Q. S.; Liang, Y. Z. Propy: A Tool to Generate Various Modes of Chou's PseAAC.

Bioinformatics **2013**, *29* (7). https://doi.org/10.1093/bioinformatics/btt072.

(33) Tsubaki, M.; Tomii, K.; Sese, J. Compound-Protein Interaction Prediction with End-to-End Learning of Neural Networks for Graphs and Sequences. *Bioinformatics* **2019**, *35* (2). https://doi.org/10.1093/bioinformatics/bty535.

(34) Axelrod, S.; Gómez-Bombarelli, R. GEOM, Energy-Annotated Molecular Conformations for Property Prediction and Molecular Generation. *Sci Data* **2022**, *9* (1). https://doi.org/10.1038/s41597-022-01288-4.

(35) Bai, P.; Miljković, F.; John, B.; Lu, H. Interpretable Bilinear Attention Network with Domain Adaptation Improves Drug–Target Prediction. *Nat Mach Intell* **2023**. https://doi.org/10.1038/s42256-022-00605-1.

(36) Shimizu, Y.; Yonezawa, T.; Bao, Y.; Sakamoto, J.; Yokogawa, M.; Furuya, T.; Osawa, M.; Ikeda, K. Applying Deep Learning to Iterative Screening of Medium-Sized Molecules for Protein–Protein Interaction-Targeted Drug Discovery. *Chemical Communications* **2023**. https://doi.org/10.1039/D3CC01283B.

(37) Guo, W.; Wisniewski, J. A.; Ji, H. Hot Spot-Based Design of Small-Molecule Inhibitors for Protein-Protein Interactions. *Bioorganic and Medicinal Chemistry Letters*. 2014. https://doi.org/10.1016/j.bmcl.2014.03.095.

(38) Lee, S.; Hu, L. Nrf2 Activation through the Inhibition of Keap1–Nrf2 Protein–Protein Interaction. *Medicinal Chemistry Research*. 2020. https://doi.org/10.1007/s00044-020-02539-y.

(39) McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J Open Source Softw* **2018**, *3* (29). https://doi.org/10.21105/joss.00861.

(40) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J Chem Inf Comput Sci* **2002**, *42* (6). https://doi.org/10.1021/ci010132r.

(41) Schreyer, A. M.; Blundell, T. USRCAT: Real-Time Ultrafast Shape Recognition with Pharmacophoric Constraints. *J Cheminform* **2012**, *4* (11). https://doi.org/10.1186/1758-2946-4-27.

(42) Allen, W. J.; Fochtman, B. C.; Balius, T. E.; Rizzo, R. C. Customizable de Novo Design Strategies for DOCK: Application to HIVgp41 and Other Therapeutic Targets. *J Comput Chem* **2017**, *38* (30). https://doi.org/10.1002/jcc.25052.

(43) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of New Features and Current Docking Performance. *J Comput Chem* **2015**, *36* (15). https://doi.org/10.1002/jcc.23905.

(44) 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.; Moore, T. W. Probing the Structural Requirements of Non-Electrophilic Naphthalene-Based Nrf2 Activators. *Eur J Med Chem* **2015**, *103*. https://doi.org/10.1016/j.ejmech.2015.08.049.

(45) Delano, W. L. The PyMOL Molecular Graphics System. *CCP4 Newsletter on protein crystallography* **2002**, *40* (1).

(46) Wang, J.; Chu, Y.; Mao, J.; Jeon, H. N.; Jin, H.; Zeb, A.; Jang, Y.; Cho, K. H.; Song, T.; No, K. T. De Novo Molecular Design with Deep Molecular Generative Models for PPI Inhibitors. *Brief Bioinform* **2022**, *23* (4). https://doi.org/10.1093/bib/bbac285.

(47) Ohue, M.; Kojima, Y.; Kosugi, T. Molecules Communication Generating Potential Protein-Protein Interaction Inhibitor Molecules Based on Physicochemical Properties. **2023**. https://doi.org/10.3390/molecules28155652.

(48) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Köpf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; Chintala, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*; 2019; Vol. 32.

(49) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; Santos Costa, A. Dos; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, [2] Alexander. Evolutionary-Scale Prediction of Atomic Level Protein Structure with a Language Model. **2021**. https://doi.org/10.1101/2022.07.20.500902.

(50) Suzek, B. E.; Wang, Y.; Huang, H.; McGarvey, P. B.; Wu, C. H. UniRef Clusters: A Comprehensive and Scalable Alternative for Improving Sequence Similarity Searches. *Bioinformatics* **2015**, *31* (6). https://doi.org/10.1093/bioinformatics/btu739.

(51) Pande, A.; Patiyal, S.; Lathwal, A.; Arora, C.; Kaur, D.; Dhall, A.; Mishra, G.; Kaur, H.; Sharma, N.; Jain, S.; Usmani, S. S.; Agrawal, P.; Kumar, R.; Kumar, V.; Raghava, G. P. S. Pfeature: A Tool for Computing Wide Range of Protein Features and Building Prediction Models. *Journal of Computational Biology* **2023**, *30* (2). https://doi.org/10.1089/cmb.2022.0241.

(52) Liu, S.; Wang, H.; Liu, W.; Lasenby, J.; Guo, H.; Tang, J. *PRE-TRAINING MOLECULAR GRAPH REPRESENTATION WITH 3D GEOMETRY*.

(53) Landrum, G. RDKit: Open-Source Cheminformatics Software, 2021.

(54) Kim, J. H.; Jun, J.; Zhang, B. T. Bilinear Attention Networks. In *Advances in Neural Information Processing Systems*; 2018; Vol. 2018-December.