

RESEARCH ARTICLE

Dual-channel graph learning reveals similarity and complementarity in protein-protein interaction networks

Tao Tang¹, Taiguang Shen¹, Weizhuo Li¹, Yangyang Chen², Sisi Yuan³, Yuansheng Liu^{4,5*}, Xinyu Yang^{4,5*}, Xiao Luo^{6*}

1 School of Modern Posts, Nanjing University of Posts and Telecommunications, Jiangsu, China, **2** Department of Computer Science, University of Tsukuba, Ibaraki, Japan, **3** School of Chinese Medicine, Hong Kong Baptist University, Kowloon, Hong Kong SAR, China, **4** College of Computer Science and Electronic Engineering, Hunan University, Hunan, China, **5** Yuelushan Laboratory, Changsha, China, **6** College of Biology, Hunan University, Hunan, China

* yuanshengliu@hnu.edu.cn (YL); xinyuyang21@hnu.edu.cn (XY); xlue@hnu.edu.cn (XL)



OPEN ACCESS

Citation: Tang T, Shen T, Li W, Chen Y, Yuan S, Liu Y, et al. (2026) Dual-channel graph learning reveals similarity and complementarity in protein-protein interaction networks. PLoS Comput Biol 22(5): e1013941. <https://doi.org/10.1371/journal.pcbi.1013941>

Editor: Belal O. Al-najjar, Al-Ahliyya Amman University, JORDAN

Received: January 23, 2026

Accepted: May 3, 2026

Published: May 20, 2026

Copyright: © 2026 Tang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All synthetic data and code necessary to reproduce the results of this study are publicly available in the github repository https://github.com/ttan6729/DMG_PPI and Zendo repository <https://zenodo.org/records/17010303>.

Abstract

Protein-protein interactions (PPIs) are governed by two fundamental interfacial mechanisms: similarity-driven, often involving symmetric structural motifs, and complementarity-driven, arising from geometric and physicochemical matching between binding surfaces. Despite their biological significance, computational models have largely overlooked the coexistence and interplay of these twofold interaction modes. Here, we introduce DMG-PPI, a dual-channel graph neural network framework that jointly models similarity and complementarity in PPI networks, extending prior heterophilous GNN concepts to explicitly disentangle these dual interaction modes. The core of the model consists of two parallel processing pathways: Alignment Message Passing (AMP), which aggregates information from proteins with similar features to capture interactions driven by shared structural patterns, and Divergence Message Passing (DMP), which emphasizes differences between proteins and identifies complementary features that may indicate physicochemical compatibility. The signals captured by AMP and DMP are integrated via an adaptive fusion strategy within each block, and the outputs of blocks are aggregated using the MixHop framework to encode higher-order interaction patterns. DMG-PPI substantially outperforms state-of-the-art methods on classical benchmark datasets, achieving a 7.19% improvement in Micro-F1 over the second-best method. Additionally, the dual-channel framework provides interpretable insights into key binding residues by identifying interfacial mechanisms. Overall, DMG-PPI serves as a powerful tool that reveals the mechanisms behind accurate PPI predictions and facilitates downstream biological analysis.

Funding: This work was supported by Yuelushan Laboratory Breeding Program (YLS-2025-ZY03024 to YL); the National Nature Science Foundation of China (62202236 to TT; 62372159 to YL, 32400506 to XL), the Natural Science Foundation of Hunan Province (2024JJ4008 to XL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Proteins interact through multiple interfacial principles rather than a single uniform mechanism, reflecting diverse modes of molecular recognition between binding partners. Accurately modeling this heterogeneity remains a central challenge in protein–protein interaction (PPI) prediction. Existing computational approaches often overlook such complexity and treat all interactions as arising from a dominant pattern. To address this limitation, we present DMG-PPI, a dual-channel graph neural network framework designed to account for distinct interaction principles within a unified model. By modeling both similarity-based and complementarity-based interaction signals, DMG-PPI enables a more comprehensive representation of interaction patterns in PPI networks. Evaluations on widely used benchmark datasets demonstrate that our approach achieves robust and consistent improvements over existing methods. Beyond prediction accuracy, DMG-PPI provides biologically meaningful interpretability by highlighting key residues and revealing distinct interfacial interaction mechanisms, offering valuable insights for downstream structural and functional studies.

Introduction

Protein–protein interactions (PPIs) constitute the structural and functional backbone of cellular systems, coordinating molecular processes ranging from signal transduction to metabolic regulation [1,2]. PPI networks thus offer a systems-level view of how molecular interactions give rise to cellular function and how their perturbation contributes to disease pathogenesis. Mapping and modeling these networks is central to both fundamental biology and biomedical research, enabling insights into molecular organization [3], complex assembly [4], system-level cellular behavior [5], target prioritization [6], and therapeutic intervention [7]. While experimental methods have produced substantial PPI data, their expense, contextual variability, and scalability constraints underscore the need for computational approaches that can predict interactions across entire networks.

At a fundamental level, protein binding is governed by two non-mutually exclusive interaction modes. **Similarity-driven interactions** arise from shared evolutionary or biochemical features that favor compatible binding affinities [8,9], whereas **complementarity-driven interactions** depend on precise geometric and physico-chemical matching at binding interfaces [10,11]. Experimental evidence supports the coexistence of these interaction modes: Yeast two-hybrid assays preferentially detect interactions between proteins with similar domain architectures [12,13], whereas X-ray crystallography reveals complementary binding at atomic resolution [14]. This duality is not unique to PPIs but recurs across biological networks, including gene regulatory [15–17], signaling [18–20], and metabolic systems [21,22], highlighting a general organizing principle of biological systems. This simplification reflects a deeper conceptual mismatch between biological interaction mechanisms and their

computational abstraction. In particular, PPI networks cannot be faithfully represented by models that conflate similarity and complementarity into a single interaction logic. Consequently, the dual nature of PPI formation has not been systematically formulated as co-equal, first-class modeling principles in computational interactomics, limiting our ability to disentangle, interpret, and reason about distinct interaction modes at the network level.

Currently, the research in computational PPI prediction primarily addresses two key aspects: learning informative representations of individual protein from their sequences or structures [23, 24], and modeling the interaction patterns within PPI networks [25]. The former aspect primarily relies on physicochemical characteristics [26,27] and residue-level encoding [28,29] of protein sequences to generate protein embeddings. More recently, Gao et al. [30] introduced graph neural networks (GNNs) to model protein tertiary structures as graphs. Graph autoencoder and masked modeling was then applied to learn a robust representation of protein structure [31]. The latter aspect concentrates on modeling interactome using GNN. Graph Isomorphism Network (GIN) was among the first GNN architectures applied to PPI prediction [32]. Subsequent studies have demonstrated improvements in generalization capabilities through Graph Attention Network (GAT) [33] and anti-symmetric Graph Convolutional Network (GCN) [34], as well as incorporating GNNs into broader frameworks [35,36].

While graph neural networks (GNNs) have become widely used for PPI prediction, most existing architectures primarily aggregate features from similar neighbors. This design inherently favors the detection of similarity-driven interactions and makes it difficult for them to recognize interactions based on complementarity. Although prior work on heterophilous GNNs has introduced mechanisms to handle structurally or functionally dissimilar nodes, these approaches were not specifically designed to disentangle the co-occurrence of similarity- and complementarity-driven interactions inherent to PPI networks. This gap constrains both predictive accuracy and biological interpretability.

Here, we introduce DMG-PPI, a Dual Message Pathway Graph Neural Network that explicitly models the dual principles governing protein–protein interactions. DMG-PPI decomposes message passing into two parallel channels: Alignment Message Passing (AMP), which captures similarity-driven interactions, and Divergence Message Passing (DMP), which captures complementarity-driven interactions through contrasting features. These channels are processed in parallel and adaptively fused to generate multi-scale representations that preserve the distinct biological semantics of each interaction mode. The AMP channel follows the homophily assumption by aggregating features from similar neighbors through summation, while the DMP channel extends heterophilous GNN principles through multiplicative message passing that explicitly contrasts node features, enabling the capture of complementary binding patterns that are biologically distinct from simple feature dissimilarity. By grounding network modeling in the physical principles of protein binding, DMG-PPI achieves superior predictive accuracy, robustness to noise, and enhanced interpretability, enabling inference of dominant binding modes and key interface features within PPI networks.

Beyond the methodological advances introduced in this work, the underlying biological question—how proteins recognize and bind to their partners—is fundamental to molecular biology and of broad interest to experimental researchers. Protein-protein interactions govern nearly all cellular processes, from signal transduction to immune responses, yet the principles that determine why certain proteins interact while others do not remain incompletely understood. By explicitly modeling two distinct interaction mechanisms—similarity-driven binding, where proteins with shared structural features associate, and complementarity-driven binding, where proteins with matching but different features interact—our framework aims to capture biological reality more faithfully than previous approaches. For experimental biologists, DMG-PPI offers practical utility beyond prediction accuracy: the model can highlight which residues are likely critical for binding (hotspots) and indicate whether a given interaction likely arises from structural similarity or physicochemical complementarity, providing testable hypotheses for mutagenesis or structural studies.

Materials and methods

Homophily is a fundamental assumption in graph theory, referring to the tendency of links to form between similar nodes. This property is prevalent in many studies on graph structure, such as social network (Fig 1A). Most previous studies of

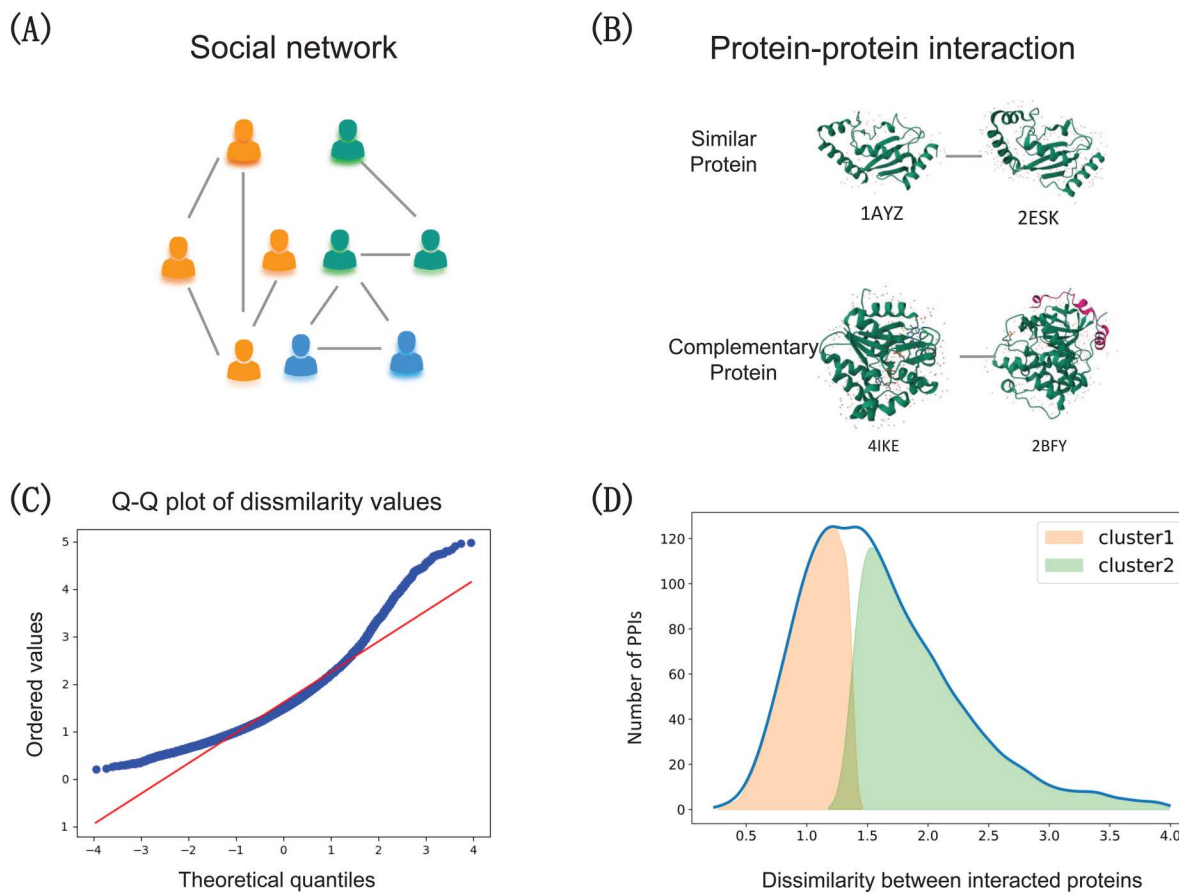


Fig 1. Protein–protein interactions (PPIs) can emerge through two principal mechanisms: similarity and complementarity. (A) In social networks, individuals with similar attributes tend to connect, a phenomenon termed homophily. (B) Analogously, in PPI network, both similar and complementary proteins tend to interact. (C) A Q-Q plot, a statistical tool for comparing distributions, confirms this pattern. The fact that many points deviate from the straight reference line indicates that the interacting protein pairs do not all follow a single, simple rule, but instead represent a mix of different interaction types. (D) Clustering analysis further categorizes the binding protein pairs into two groups exhibiting markedly different levels of structural dissimilarity. These groups correspond to interactions primarily driven by structural similarity and complementarity, respectively.

<https://doi.org/10.1371/journal.pcbi.1013941.g001>

PPI networks have largely relied on this assumption, favoring GNN architectures with sum-based aggregation that naturally emphasize homophilous relationships. However, as illustrated in Fig 1B, biological evidence indicates that protein interactions often arise not only between structurally or functionally similar proteins but also between complementary counterparts, reflecting the coexistence of similarity- and complementarity-driven interactions.

To systematically investigate protein interaction patterns, we quantified the similarity between interacting proteins as the euclidean distance between their embeddings derived from pre-trained geometric model [31]. A Q–Q plot of similarity scores for interacting pairs in SHS27K is shown in Fig 1C, where numerous points deviate substantially from the reference line, suggesting a mixed distribution of similarity among interacting proteins. Subsequent clustering analysis (Fig 1D) further supports the coexistence of multiple connection patterns in PPI networks. Specifically, we computed pairwise similarity scores for interacting protein pairs and examined their empirical distribution. The resulting distribution exhibits two distinct peaks, forming a clear bimodal pattern. This bimodality suggests that the interaction samples naturally partition into two separable groups: one characterized by high similarity, consistent with similarity-driven interactions, and the other characterized by lower similarity, indicative of complementarity-driven relationships. The presence of two stable density

modes indicates that protein interactions cannot be explained by a single uniform mechanism, but instead arise from at least two major binning principles. Together, these findings highlight the complex nature of PPIs, in which similarity-based and complementary mechanisms jointly contribute to network connectivity.

Here, we propose DMG-PPI for modeling the two-fold interaction patterns in PPI network. As depicted in Fig 2, our approach comprises four main modules: (1) construction of the physical protein graph, (2) a graph autoencoder for data preprocessing and generation of initial node embeddings, (3) a dual-channel GNN framework for capturing complex interaction patterns within the PPI network, and (4) a classifier for final prediction.

Problem formulation

The PPI network is denoted as $\mathcal{G} = (\mathcal{P}, \mathcal{I})$, where $\mathcal{P} = \{p_1, p_2, \dots, p_n\}$ is the protein set, and $\mathcal{I} = (p_i, p_j, y_{ij})$, the protein feature of p_i is defined as x_i . The objective is to learn a model $(p_i, p_j) \rightarrow \hat{y}_{ij}$ from training set.

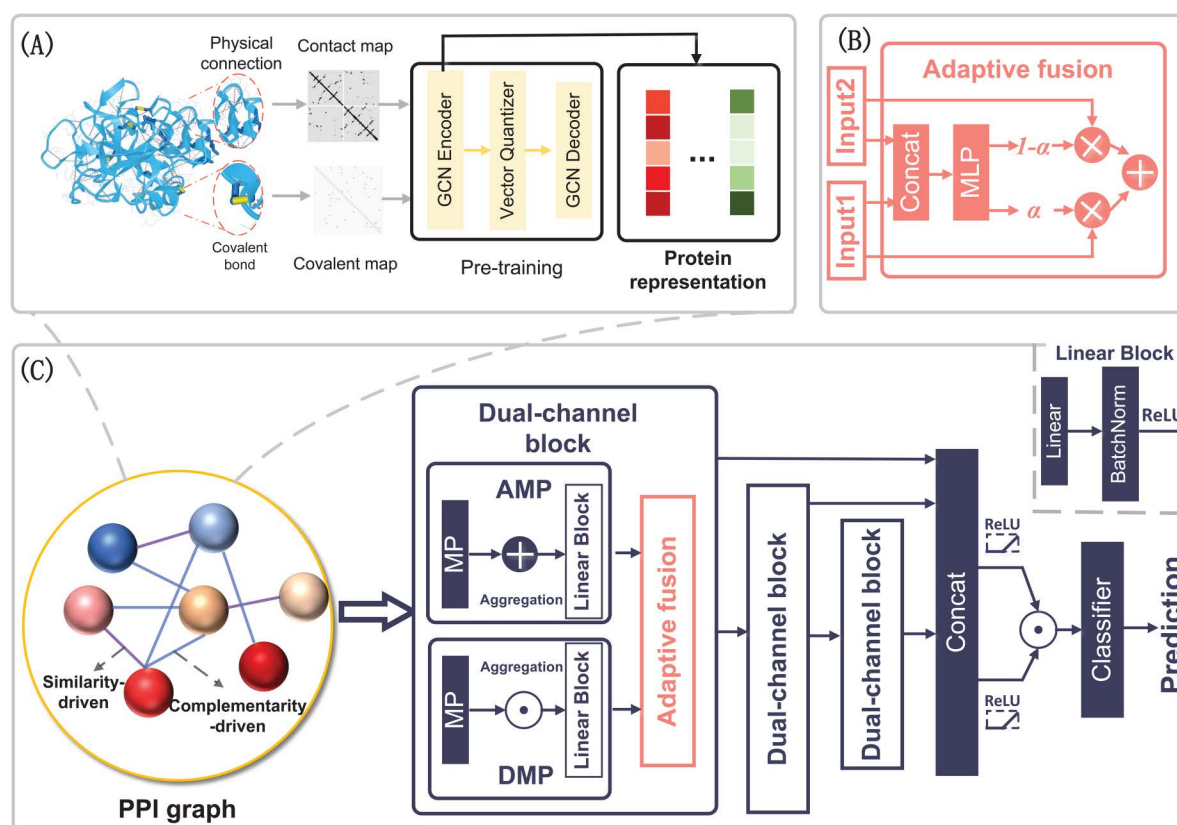


Fig 2. (A) Each protein is first converted into a structural representation derived from its contact map and covalent connectivity. These structural features are encoded into compact protein embeddings using a pre-trained graph neural network autoencoder. In simple terms, this step transforms complex structural information into a learnable numerical representation of each protein. **(B) The adaptive fusion block integrates different sources of structural information using a gating mechanism, which automatically determines how much weight to assign to each feature type depending on the context.** **(C) The core of DMG-PPI models two complementary principles of protein interactions.** AMP captures similarity-based relationships, where proteins with similar structural patterns tend to interact. DMP captures complementarity-based relationships, where structurally distinct but functionally compatible proteins interact. The outputs of these channels are combined to form refined protein representations. Finally, the interaction between two proteins is quantified by combining their representations into a pair-level descriptor, which is used to predict whether the proteins interact.

<https://doi.org/10.1371/journal.pcbi.1013941.g002>

Feature extraction from intramolecular physical and chemical connections

We use residue-level graph to model the tertiary structure of each protein. Unlike the previous methods that only learn the spatial contact inside protein, our approach considers both the physical and chemical connections. For each protein, two residual-level protein graphs are constructed to model the physical and chemical connections respectively, both graphs consist of an adjacency matrix $A_p \in \{0, 1\}^{r \times r}$ and a feature matrix $M_p \in \mathbb{R}^{r \times s}$, where r denotes the residue number of protein.

For chemical protein graph, the adjacency matrix is derived from covalent bonds between atoms; if there is a covalent bond between two atoms, then there is a connection between the corresponding residues. Covalent bonds are identified based on known chemical bonding patterns within amino acid residues and ligand molecules. For each pair of atoms, the connection is determined by the presence of a covalent bond as defined by standard residue chemistry or explicit bonding information in the structure data. The node feature is formed by three physicochemical properties: solvent accessible surface area (SASA), electrostatic potential surface (EPS), and pKa shifts. For the physical connection protein graph, a contact map is used as the adjacency matrix, in which the connection between residues is determined by the 3D coordinates of alpha carbon atom in residue. For each pair of residue, the connection exists if the distance between their alpha carbon atoms is smaller than 8Å. The node features are extracted by the type-dependent encoding of RDKit [37].

All computed node features are subsequently normalized across the dataset prior to model input to ensure numerical stability and comparability across proteins. The physical and chemical protein graphs are processed with graph auto-encoder to learn the embeddings of protein graph, we adapt the masked vector quantizer [31] for training process. The protein graph are processed by GNN encoder, GNN decoder and a vector quantizer:

$$\begin{aligned} Z &= \text{GNN}_{\text{enc}}(M_p, A_p) \\ Q &= \text{VQ}(Z) \\ \hat{M}_p &= \text{GNN}_{\text{dec}}(Z, Q), \end{aligned} \quad (1)$$

the masked modeling m is applied on output of vector quantizer:

$$\begin{aligned} Q_{\text{mask}} &= m \odot Q, m \in \{0, 1\}^d \\ \hat{M}_{\text{recon}} &= \text{GNN}_{\text{dec}}(Z, Q_{\text{mask}}). \end{aligned} \quad (2)$$

The graph autoencoder is pre-trained on SHS148K. The reconstruction loss is defined as the sum of three components: the mean squared error (MSE) between quantized representations Q and latent representation Z ; the MSE between the original feature M_p and its reconstruction $\hat{M}_p$; and the MSE between original feature M_p and masked reconstruction $\hat{M}_{\text{recon}}$.

The embedding of each protein h_i is obtained by concatenating the latent representations from the physical and chemical graphs, followed by a readout operation: $h_i = \text{READOUT}(Z_{\text{phy}} \| Z_{\text{che}})$, $\|$ denotes the concatenation operation.

Dual-channel framework for learning interaction pattern

DMG-PPI employs two parallel channels to jointly model similarity- and complementarity-driven interactions simultaneously. Given a PPI graph $\mathcal{G} = (\mathcal{P}, \mathcal{I})$, along with the initial node embeddings $\mathbf{h}$ for each protein, DMG-PPI propagates information through a stack of dual-channel GNN blocks, $\mathbf{h}_i^{t+1}$ denotes the output of t th block. Within each block, protein representations are independently processed through two specialized message-passing pathways: AMP, designed to capture similarity-driven interactions, and DMP, tailored for complementarity-driven interactions. The outputs of AMP and DMP are formally defined as:

$$\begin{aligned} \mathbf{h}_{i,\text{AMP}}^{(l+1)} &= \sigma \left(\sum_{j \in \mathcal{N}(i) \cup \{i\}} \frac{1}{\sqrt{d_i d_j}} \mathbf{W}_{\text{AMP}}^{(l)} \mathbf{h}_j^{(l)} \right) \\ \mathbf{h}_{i,\text{DMP}}^{(l+1)} &= \sigma \left(\bigodot_{j \in \mathcal{N}(i) \cup \{i\}} \mathbf{W}_{\text{DMP}}^{(l)} \mathbf{h}_j^{(l)} \right), \end{aligned} \quad (3)$$

where $\mathbf{W}_{\text{AMP}}^{(l)}$ and $\mathbf{W}_{\text{DMP}}^{(l)}$ denote the learnable weight matrix of each channel, $\mathcal{N}(i)$ denotes the neighborhoods of p_i , d_i denotes the degree of the i th protein. AMP and DMP employ additive and multiplicative message-passing pathways, respectively, to capture the distinct biological characteristics of PPI networks. AMP aggregates features from neighboring proteins through summation, which smooths the representations and naturally reinforces similarity-driven interactions. In contrast, DMP emphasizes physicochemical complementarity between interacting proteins. By combining neighbor features multiplicatively, DMP amplifies contributions from proteins with complementary structural or functional characteristics, reflecting the underlying biophysical principle that complementary surfaces or motifs facilitate stable interactions.

A gating mechanism is introduced to adaptively balance the contribution of both channels. The gate coefficient $\alpha \in (0, 1)$ is obtained by linearly projecting the concatenated AMP and DMP embeddings:

$$\alpha = \sigma \left(\mathbf{W}_g \cdot [\mathbf{h}_{i,\text{AMP}}^{(l+1)} \parallel \mathbf{h}_{i,\text{DMP}}^{(l+1)}] + b_g \right), \quad (4)$$

where $\mathbf{W}_g$ and b_g are learnable parameters, σ denotes an activation function that constrains α in range $(0, 1)$. In this study, we employ sigmoid function. This coefficient adaptively regulates the relative contribution of each channel to the fused representation, as follows:

$$\mathbf{h}_i^{(l+1)} = \sigma(\text{BN}(\alpha \cdot \mathbf{h}_{i,\text{AMP}}^{(l+1)} + (1 - \alpha) \cdot \mathbf{h}_{i,\text{DMP}}^{(l+1)})). \quad (5)$$

To capture multi-scale dependencies and alleviate the over-smoothing effects commonly observed in deep GNNs, we adopt a MixHop-based aggregation scheme for computing the final protein representation. In the MixHop framework, the embedding of each block is concatenated together, followed by linear transformation, batch normalization and activation function sequentially. The final node embedding is as following:

$$\mathbf{x}_i = \sigma(\text{BN}(\mathbf{W}_c \cdot [\mathbf{h}_i^1 \parallel \mathbf{h}_i^2 \parallel \dots \parallel \mathbf{h}_i^l] + b_c)). \quad (6)$$

Classifier and training

For prediction, the representation of candidate protein pair (p_i, p_j) is derived by element-wise multiplication of their embeddings, followed by a three-layer multilayer perceptron (MLP):

$$\hat{y}_{ij} = \text{MLP}(\mathbf{x}_i \odot \mathbf{x}_j). \quad (7)$$

The task is formulated as a multi-label classification problem, in which each pair of proteins may participate in multiple types of PPIs. We utilize a multi-task binary cross-entropy loss for training, defined as:

$$\mathcal{L} = \sum_{x_{ij} \in \mathcal{X}_{\text{train}}} \frac{1}{C} \sum_{k=1}^C \ell_{\text{BCE}}(y_{ij}^k, \hat{y}_{ij}^k), \quad (8)$$

where C is the number of PPI types, $y_{ij}^k \in \{0, 1\}$ denotes the ground-truth label for type k , $\hat{y}_{ij}^k$ is the predicted probability, and $\ell_{\text{BCE}}(y, \hat{y}) = -y \log \hat{y} - (1 - y) \log(1 - \hat{y})$.

Results

We benchmark DMG-PPI on SHS27K and SHS148K [30,33,38,39], two canonical multi-type PPI datasets curated from STRING [40], a comprehensive resource for human interactome analysis. Both datasets were filtered to retain only high-confidence interactions (score ≥ 0.7), resulting in 12,517 PPIs in SHS27K and 44,488 PPIs in SHS148K. To minimize protein overlap between the training and test sets, we adopt the Breadth-First Search (BFS) and Depth-First Search (DFS) splitting strategies proposed by Lv et al. [39]. During the experiments, DMG-PPI was trained with a learning rate of 1×10^{-3} and a weight decay of 5×10^{-4} . The number of layers was set to three, and the hidden dimension was fixed at 256, without applying dropout. Training was conducted using mini-batch gradient descent, and early stopping was employed based on the validation loss. Specifically, training was terminated if no improvement was observed for 20 consecutive epochs.

The benchmark methods include PIPR [41] (Siamese residual recurrent convolutional network), ESM2_AMP [42] (Uses the ESM2 protein language model to extract sequence segment feature), TUNA [43] (Transformer-based Uncertainty Aware model for PPI), GNN_PPI [32] (1d convolution and graph convolution), AFTGAN [33] (attention-free transformer and graph attention mechanisms), LDMGNN [44] (Graph neural network-based method for multi-graph learning of PPIs), HIGH-PPI [30] (Hierarchical PPI prediction model that leverages multi-level graph information to capture interaction topology), BaPPI [34] (Anti-symmetric GNN framework for balanced learning), and MAPE-PPI [31] (Microenvironment-aware protein embedding model that encodes residue environments with structural context).

DMG-PPI achieves the best performance and model capabilities

Fig 3A presents the results of DMG-PPI compared with leading benchmark methods. Each experiment was repeated five times, with mean values (two decimal places) reported; standard deviations are provided in S1–S4 Tables. Prediction performance is quantified by: (i) Micro-F1: the harmonic mean of precision and recall; (ii) AUPRC: area under the precision–recall curve; (iii) NDCG: normalized discounted cumulative gain; (iv) P@500L: proportion of positive PPIs among the top 500 ranked links; (v) AUROC: area under the receiver operating characteristic curve; and (vi) z-score aggregated across the above metrics.

DMG-PPI consistently outperforms all state-of-the-art methods across every metric. MAPE-PPI and BaPPI achieve competitive second-place performance in different scenarios, while PIPR demonstrates notably inferior performance, likely due to its non-GNN architecture, which limits its ability to capture global topological information. Given the high class imbalance and skewed link distribution in PPI datasets, Micro-F1 provides a more reliable performance measure than NDCG or AUROC, as the latter tend to overestimate performance when generalization is weak [45]. This explains the smaller differences observed between methods under NDCG and AUROC. Quantitatively, DMG-PPI exceeds the second-best method in Micro-F1 by 7.19% on SHS27K and 3.14% on SHS148K.

We further evaluate robustness by introducing label perturbations into the training set to simulate undiscovered PPIs, which is a common feature of PPI datasets. Positive labels for each interaction were randomly removed at ratios ranging from 10–50% to simulate the frequently occurring undiscovered interactions. As shown in Fig 3C, DMG-PPI sustains an F1-score above 0.7 even with 30% label removal, with negligible performance loss when the ratio is below 30%. This resilience indicates that DMG-PPI captures stable global interaction patterns that persist despite incomplete supervision, particularly in high-recall regimes. At low recall, the model trained with complete labels retains a clearer advantage, underscoring robustness in top-ranked predictions.

Notably, the superior performance of DMG-PPI can be attributed to its enhanced capability to recognize positive PPIs from the minority class. To illustrate this discrepancy, we examine four representative positive PPI cases formed

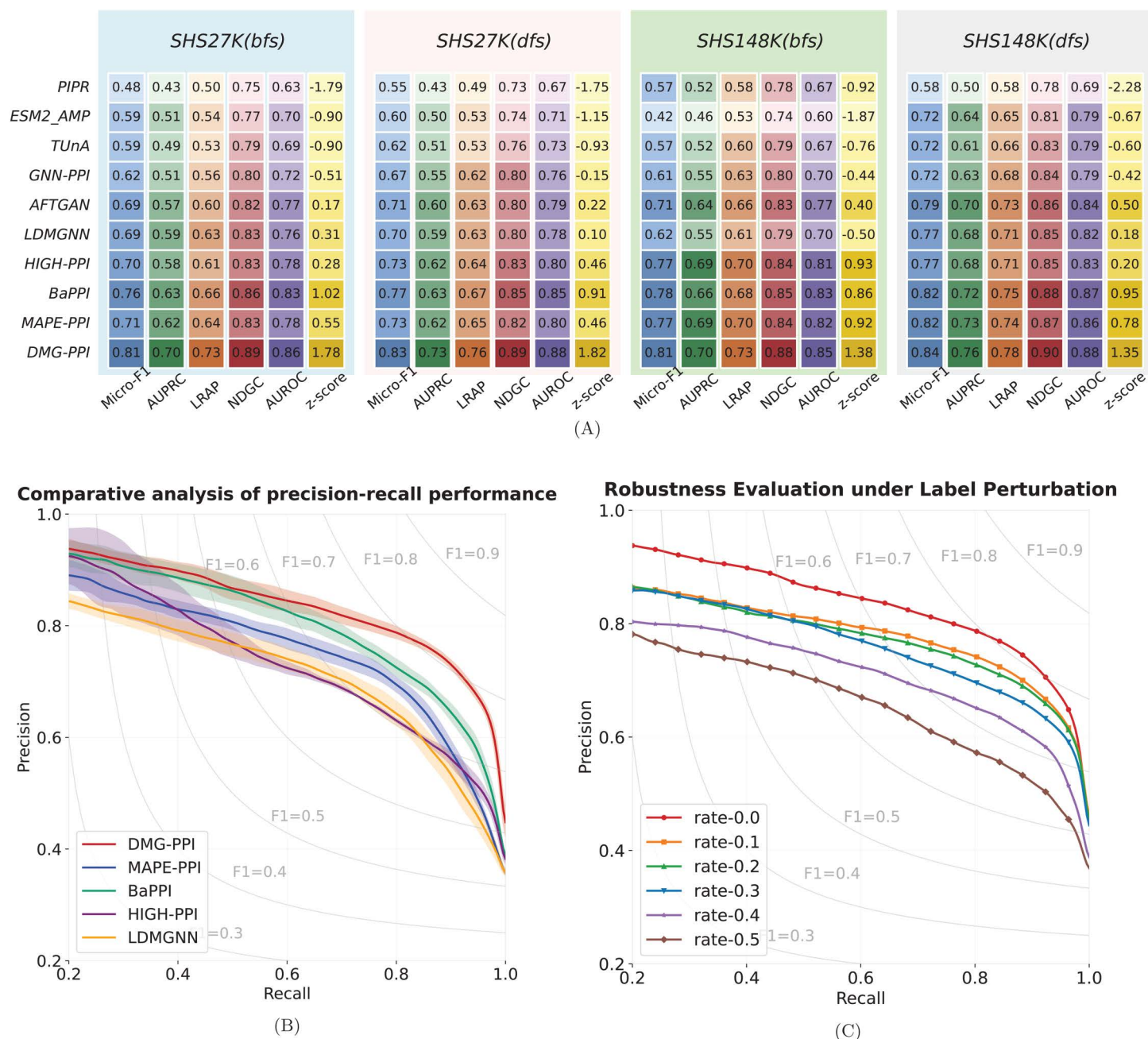


Fig 3. Comparative performance of DMG-PPI and benchmark methods. (A) Heatmaps showing the performance of each method across the following evaluation metrics: Micro-F1, AUPRC, LRAP, NDCG, AUROC, and overall z-score. Darker colors indicate better performance. Each row indicates the performance of corresponding method. (B) Precision–recall curves for PPI prediction on SHS27K, comparing the top five methods; shaded areas represent the range between the highest and lowest results. (C) Robustness of DMG-PPI to label perturbation on SHS27K, evaluated using precision–recall curves at different perturbation ratios.

<https://doi.org/10.1371/journal.pcbi.1013941.g003>

by biologically important proteins. As shown in Fig 4, DMG-PPI assigns systematically higher confidence scores to true positive interactions, with the minimum predicted probability exceeding 0.96. In contrast, the predicted probabilities produced by TUnA and HIGH-PPI for the same positive PPIs fall within a much lower range of 0.20 to 0.53, with most values concentrated between 0.3 and 0.4. DMG-PPI exhibits a clear separation between positive and negative predictions, this pronounced separation in predicted probabilities highlights its superior discriminative power and reliability in identifying minority positive PPIs.

DMG-PPI precisely detects key interface residues underlying PPIs

Hotspots are specific subsets of residues that contribute disproportionately to the binding free energy, hence play a fundamental role in governing protein–protein interactions. The divergence message pathway (DMP) focuses on physicochemical and geometric complementarity at the protein interface, which are key factors underlying the formation of hotspots. The residue-level DMP scores allow estimation of the probability that a residue serves as such a site.

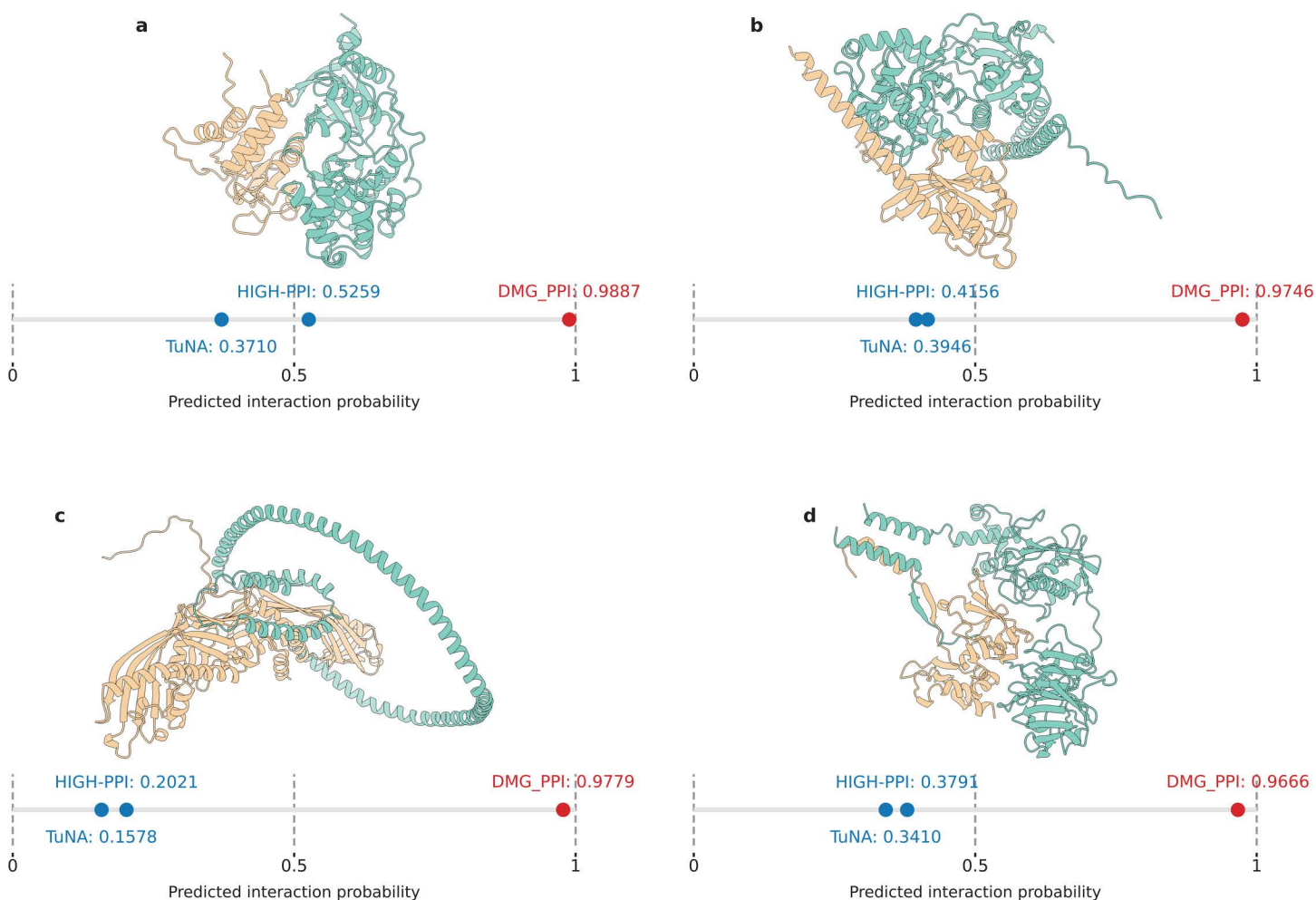


Fig 4. Protein–protein interaction (PPI) examples that are correctly predicted by DMG-PPI but not by HIGH-PPI or TUnA. The corresponding PPI complex structures are predicted using GRAMM and visualized with ChimeraX. **(A)** O15264 (Mitogen-activated protein kinase 13) and P51452 (Dual specificity protein phosphatase 3); **(B)** P11233 (Ras-related protein Ral-A) and Q15208 (Serine/threonine-protein kinase 38); **(C)** P11597 (Cholesteryl ester transfer protein) and P02647 (Apolipoprotein A-I); **(D)** P01033 (Metalloproteinase inhibitor 1) and Q9H306 (Matrix metalloproteinase-27).

<https://doi.org/10.1371/journal.pcbi.1013941.g004>

We systematically mask each residue individually and compute the resulting percentage change in both DMP and the prediction. A normalized Bayesian-style product of these two changes is then used to quantify the likelihood of each residue being a hotspot. We provide a case study of hotspot probability of interaction between HMBOX1 (UniProt ID O76528) and CSK (UniProt ID P41240). The proxy ground truth is computed by Rosetta Flex ddG. As shown in Fig 5A, the tendency of being hotspot is represented by color, DMG-PPI identifies nearly all hotspots with high accuracy. In particular, it successfully detects hotspots with low SASA (e.g., residue 296 of CSK, 3.76 Å²), which exhibit limited structural exposure and are often overlooked by conventional methods that rely solely on surface geometry. The ability of DMG-PPI to recognize these less apparent, structurally buried hotspots underscores its broader applicability and robustness in protein–protein interaction analysis.

Furthermore, since the two channels of DMG-PPI correspond to shape complementarity and structural motifs, respectively, their associated gating coefficients can be leveraged to infer the dominant driving force of PPI. We provide a case study of the driving force that relates two proteins: DVL2, a well-connected, high-centrality node in Wnt signaling networks [46,47], and ARF5, an essential functional protein [48] involved in vesicular trafficking, the ground truth is obtained from IntAct [49]. The confidence score of driving force is derived from the gating coefficients in our dual-channel framework. As shown in Fig 5B, experimentally verified driving force is represented by solid line, while prediction is indicated by dashed line, the color of edges demonstrate the confidence score of driving force. DMG-PPI correctly predicts the main driving force of all PPIs with available ground truth, with particularly strong accuracy observed for interactions such as DVL2–LPR6, DVL2–AXIN1, and ARF5–COPB. The accurate prediction provided by DMG-PPI can relieve the high cost of experimental method, the excellent interpretability also demonstrates DMG-PPI learns generalized protein representation.

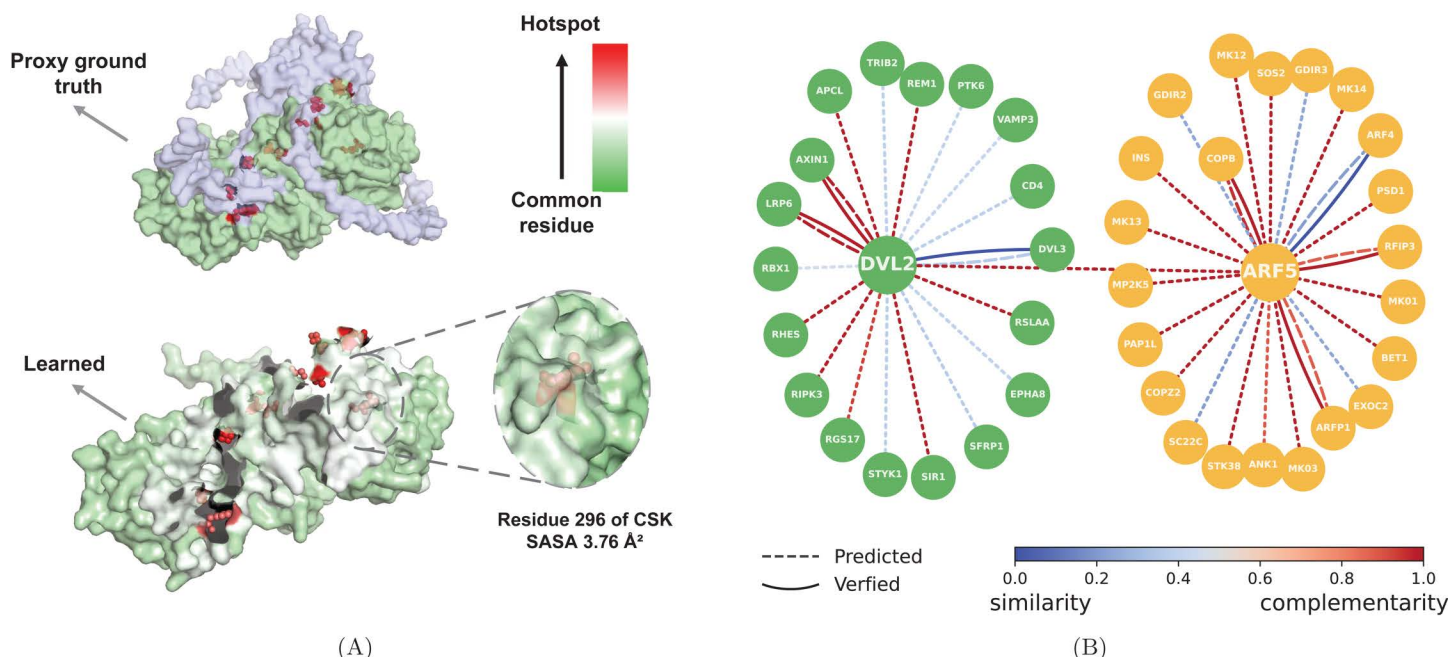


Fig 5. The interpretability provided by DMG-PPI. (A) Surface representation of a protein complex (UniProt IDs: O76528 and P41240). The likelihood of a residue being a hotspot is visualized using a color gradient from low (green) to high (red). (B) DMG-PPI accurately identifies the interaction mode between proteins. Dashed lines represent the prediction computed by DMG-PPI, while solid lines represent ground-truth. The modes of PPIs are represented by edge color, varying along a spectrum from blue (similarity-driven) to red (complementarity-driven).

<https://doi.org/10.1371/journal.pcbi.1013941.g005>

Learning dual interaction patterns enhances performance

The key innovation of our approach lies in a dual-channel framework that integrates distinct message-passing pathways, each tailored to a specific interaction pattern in the PPI network. To assess the contribution of learning dual interaction patterns, we compare the performance of the full dual-channel model with its single-channel variants, where only the AMP or DMP is used. The evaluation is conducted from the following perspectives: (1) an ablation study examining the individual contribution of each channel, (2) a type-specific performance analysis in terms of AUPR, and (3) an analysis of the interaction patterns among the top 500 predicted PPIs.

We first examine the performance of each channel on SHS27K and SHS148K. As illustrated in Fig 6A, the dual-channel framework substantially outperforms each individual message pathway on both datasets. These results indicate that each message pathway is limited to learning only a portion of the structural information inherent in the PPI network, thereby verifying the coexistence of interaction patterns and the effectiveness of our approach. Moreover, the variance of dual-channel framework is considerably lower than that of AMP and DMP, highlighting the model's ability to achieve balanced learning across the two interaction patterns and its strong robustness. Notably, AMP performs better on SHS148K, whereas DMP yields more accurate predictions on SHS27K. This discrepancy may be attributed to the higher similarity among interacting proteins in SHS148K compared to SHS27K, which makes similarity-driven interactions more prominent in the former.

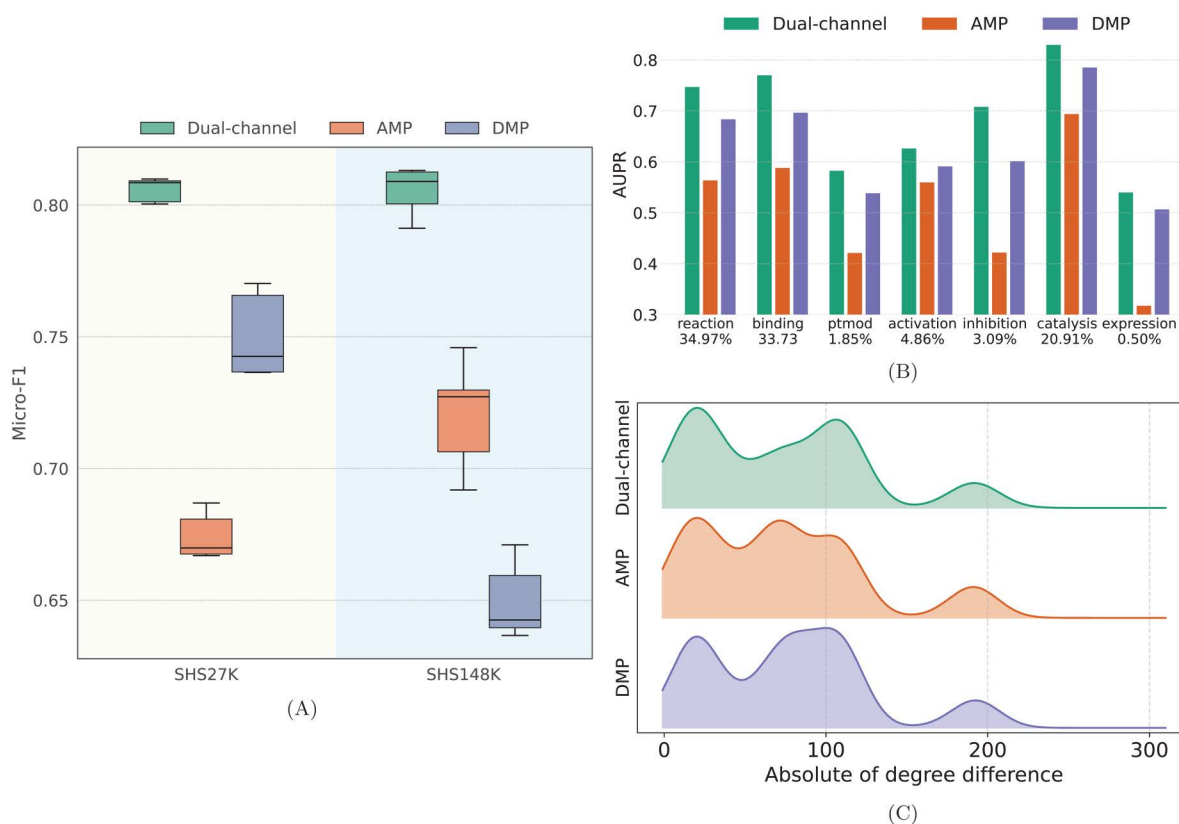


Fig 6. Performance comparison between each message pathway. (A) The Micro-F1 of each variant on SHS27K and SHS148K. (B) The AUPR across PPI types on SHS27K, the x-axis labels denote PPI types and the corresponding percentage. (C) Patterns of top-500 PPIs predicted by each variant. The x-axis represents the absolute value of the degree difference between the interacting proteins, while the y-axis indicates the percentage.

<https://doi.org/10.1371/journal.pcbi.1013941.g006>

We further evaluate each channel's performance across interaction types on the SHS27K dataset. In PPI networks, the distribution of interaction types is highly imbalanced: the dominant types (reaction, binding, and catalysis) account for over 80% of all interactions. Such skewed distribution may bias prediction models, making balanced performance across types a meaningful indicator of generalization. As shown in Fig 6B, the dual-channel framework significantly outperforms AMP and DMP across all interaction types. The performance gains of the dual-channel over AMP are especially pronounced for binding, ptmod, inhibition, and expression types that are strongly associated with biochemical complementarity (e.g., charge, shape, and hydrophobicity). These results demonstrate that our approach effectively addresses the gap in learning complementarity patterns within PPI networks. Particularly, ptmod interactions hold considerable biological significance, as they are central to protein function regulation, signal transduction pathways, and disease mechanisms. Improvements in predicting these interactions may thus have broad implications for downstream biomedical applications, including disease gene prioritization and drug target identification.

To examine whether the learned gating coefficient reflects similarity-driven interaction patterns, we computed the Spearman correlation between α and sequence identity for binding interactions in SHS148K. The analysis yielded a moderate positive correlation ($\rho = 0.377$, $p = 0.071$), indicating a trend in which higher sequence similarity corresponds to increased weighting of the similarity-driven (AMP) pathway. Although not statistically significant at the 0.05 level, this result supports the interpretation of α as capturing interaction mode preference.

Finally, to probe the interaction patterns captured by each pathway, we analyze the top 500 predicted PPIs from the dual-channel, AMP, and DMP models on SHS27K. Specifically, we compare the absolute degree difference between interacting protein pairs. As illustrated in Fig 6C, certain recurring patterns emerge, such as peaks around degrees 25, 100, and 175. The distribution associated with the dual-channel model is more densely concentrated than those of AMP and DMP, suggesting that it captures more distinct structural characteristics and integrates them more effectively.

DMG-PPI reveals the relationship between PPI types

In applications such as mechanistic modeling and pathway reconstruction [50], information about individual PPI types is often insufficient. Investigating the relationships between different PPI types can yield deeper mechanistic insights into molecular functions and disease mechanisms. To explore these inter-type functional dependencies, we decompose the classifier's weight matrix into a set of $k \times k$ vectors based on the corresponding PPI type. The L2 norm of each vector is used to quantify the influence between PPI types.

The inferred influence map produced by DMG-PPI is presented in Fig 7, where each row indicates how all PPI types contribute to a specific PPI type. Notably, the inferred relationships align with some known biological mechanisms. For example, the reaction type is most influenced by ptmod and catalysis, consistent with previous findings that reaction events frequently occur in catalytic or modifying contexts [51,52]. The matrix is notably asymmetric, revealing directional dependencies rather than reciprocal relationships. For instance, ptmod strongly influences reaction, whereas the reverse effect is comparatively weaker, consistent with their mechanistic hierarchy. Quantitatively, reaction, binding, ptmod, and catalysis exhibit the highest aggregated incoming influence scores, indicating that these types are more context-dependent.

DMG-PPI learns topology-robust representations beyond local motif density

To assess whether DMG-PPI learns unbiased representations that are not strongly influenced by local network topology, we compared embedding distance distributions for edges located in motif-dense versus motif-poor neighborhoods. The motif density of each edge was quantified as the sum of L3 motifs associated with its two endpoints. Edges in the bottom 20% and top 20% of motif density were defined as motif-poor and motif-richness, respectively. As shown in Fig 8, the resulting embedding distance distributions largely overlap, with only minor differences observed between the two groups. These findings suggest that DMG-PPI does not preferentially encode interactions in dense topological motifs, but instead

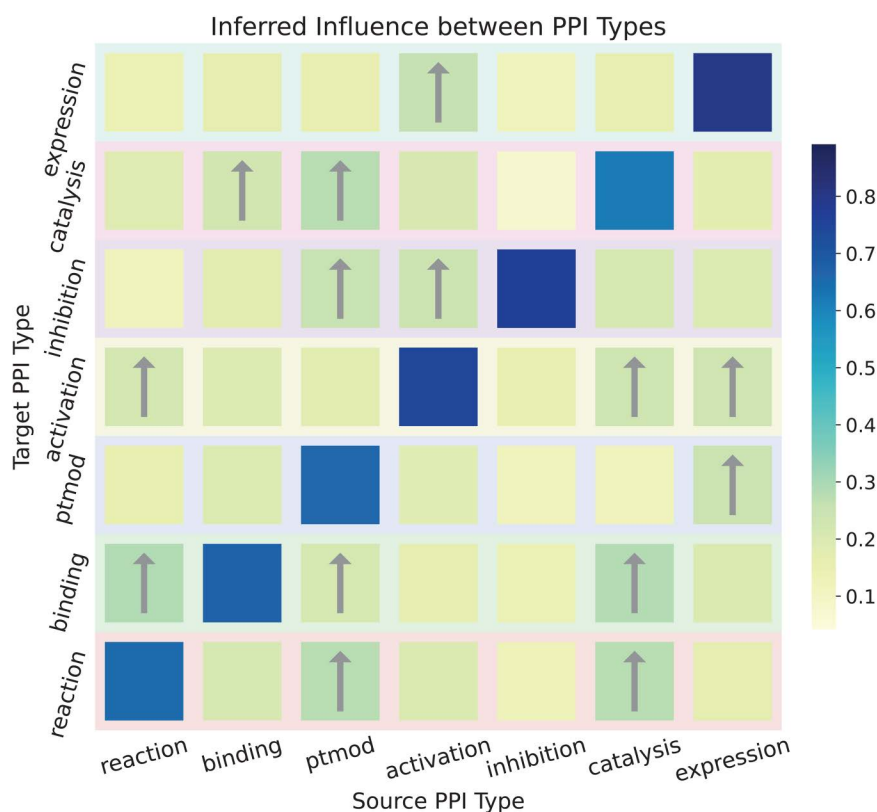


Fig 7. Directional influence among PPI types. Each row corresponds to a specific PPI type, while each column represents the type influencing it. The value at row i and column j indicates how strongly type j contributes to type i , as learned by DMG-PPI. Arrows mark influences stronger than 0.2. Overall, the map highlights directional relationships among interaction types and reflects known biological mechanisms.

<https://doi.org/10.1371/journal.pcbi.1013941.g007>

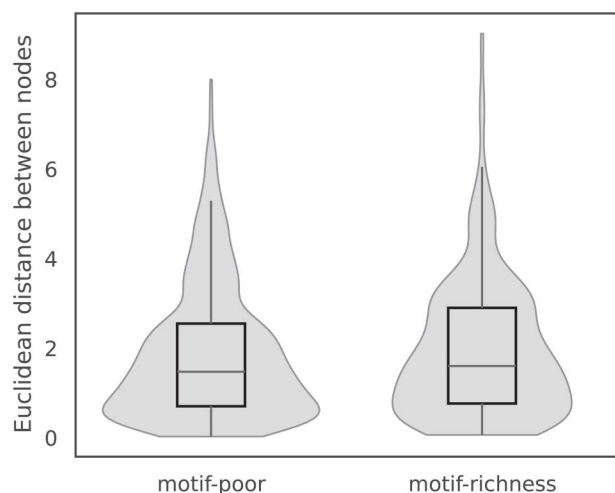


Fig 8. Distribution of Euclidean distances between node embeddings for edges in motif-richness versus motif-poor regions. The similarity of distances across regions suggests that DMG-PPI learns representations largely independent of local motif density.

<https://doi.org/10.1371/journal.pcbi.1013941.g008>

learns representations that are relatively insensitive to local motif overrepresentation, consistent with a balanced integration of network context and protein-specific features.

Further, linear regression analysis indicates that motif richness explains only 1.6% of the variance in embedding distances ($R^2=0.016$), suggesting that DMG-PPI embeddings are largely independent of local motif density. Consistently, Spearman correlation analysis reveals a weak monotonic relationship between motif richness and embedding distance ($\rho = 0.10$, $p = 1.72 \times 10^{-8}$), highlighting that the relative organization of edges in the embedding space is minimally influenced by motif structure. Together, these results demonstrate that DMG-PPI captures topology-robust representations that extend beyond the influence of local motif density.

Discussion

In this paper, we introduce DMG-PPI, an efficient PPI prediction method that models the complex interaction patterns in PPI networks from a dual-channel GNN perspective. In the preprocessing stage, the physical connection and covalent bond are extracted as initial protein representation. At interactome level, we employ two channels that use additive and multiplicative aggregation to capture complementarity-driven and similarity-driven interactions, respectively. Experiments on classical benchmark datasets demonstrate that DMG-PPI significantly outperforms state-of-the-art methods in PPI prediction. Our framework also proves to be highly reliable, performing well even on rarer types of PPIs and when the input data is incomplete or noisy. Most importantly for biological discovery, DMG-PPI offers clear interpretability: by analyzing its internal calculations, we can identify whether an interaction is likely driven by similarity or complementarity and pinpoint the specific residues involved, providing concrete, testable hypotheses for experimental follow-up. Overall, the dual-channel framework provides a powerful tool for systematic characterization of interactome connectivity and enhances mechanistic insights into PPIs.

While DMG-PPI achieves promising performance, some areas are worth consideration for further improvements. (1) This study primarily investigated the driving forces of PPIs in the context of binary interactions. However, in multiprotein complexes, the overall stability and functionality often emerge from a mosaic of interaction modes across different protein pairs. Modeling these interdependencies could provide a more faithful representation of the biophysical principles governing complex formation. (2) The potential of a unified protein graph that simultaneously incorporates physical and chemical features remains unexplored. A geometric GNN that directly operates on the full 3D atomic coordinates—integrating shape, electrostatics, and residue-level properties into a single graph could enable the learning of physically interpretable embeddings, potentially improving generalization to unseen proteins. (3) Although DMG-PPI employs a dual-channel GNN to capture both similarity- and complementarity-driven interactions, its effectiveness still depends on the availability of informative topological signals in the interaction network. For proteins located in relatively sparse or atypical regions of the network, the limited neighborhood information may reduce the effectiveness of graph propagation, potentially leading to missed or low-confidence predictions for certain interactions. (4) In addition, current multi-label formulation treats non-annotated interaction types of known protein pairs as negative labels and does not incorporate non-interacting protein pairs. Given the incompleteness of PPI annotations, this assumption may introduce potential bias. Developing more refined negative sampling or confidence-aware labeling strategies may further enhance robustness and biological validity.

Supporting information

S1 Table. Detailed performance comparison of DMG-PPI and benchmark methods on SHS27K under the BFS scheme.

(PDF)

S2 Table. Detailed performance comparison of DMG-PPI and benchmark methods on SHS27K under the DFS scheme.

(PDF)

S3 Table. Detailed performance comparison of DMG-PPI and benchmark methods on SHS148K under the BFS scheme.

(PDF)

S4 Table. Detailed performance comparison of DMG-PPI and benchmark methods on SHS148K under the DFS scheme.

(PDF)

S1 Fig. Protein–protein interaction between P04004 and P05231 that is not correctly identified by existing methods.

(PDF)

Author contributions

Conceptualization: Tao Tang, Taiguang Shen, Weizhuo Li, Sisi Yuan, Yuansheng Liu, Xinyu Yang, Xiao Luo.

Data curation: Tao Tang, Taiguang Shen, Sisi Yuan.

Formal analysis: Taiguang Shen, Weizhuo Li, Yuansheng Liu, Xiao Luo.

Investigation: Tao Tang, Xinyu Yang.

Methodology: Tao Tang, Taiguang Shen.

Resources: Yangyang Chen.

Software: Tao Tang.

Validation: Tao Tang, Yangyang Chen, Xiao Luo.

Visualization: Tao Tang, Yangyang Chen, Xiao Luo.

Writing – original draft: Tao Tang.

Writing – review & editing: Yuansheng Liu.

References

1. Thakur AK, Movileanu L. Real-time measurement of protein-protein interactions at single-molecule resolution using a biological nanopore. *Nat Biotechnol.* 2018;37(1):96–101. <https://doi.org/10.1038/nbt.4316> PMID: 30531896
2. Wang X-W, Madeddu L, Spirohn K, Martini L, Fazzone A, Becchetti L, et al. Assessment of community efforts to advance network-based prediction of protein-protein interactions. *Nat Commun.* 2023;14(1):1582. <https://doi.org/10.1038/s41467-023-37079-7> PMID: 36949045
3. Lu H, Zhou Q, He J, Jiang Z, Peng C, Tong R, et al. Recent advances in the development of protein-protein interactions modulators: mechanisms and clinical trials. *Signal Transduct Target Ther.* 2020;5(1):213. <https://doi.org/10.1038/s41392-020-00315-3> PMID: 32968059
4. Nussinov R, Tsai C-J, Jang H. Protein ensembles link genotype to phenotype. *PLoS Comput Biol.* 2019;15(6):e1006648. <https://doi.org/10.1371/journal.pcbi.1006648> PMID: 31220071
5. Lee AC-L, Harris JL, Khanna KK, Hong J-H. A comprehensive review on current advances in peptide drug development and design. *Int J Mol Sci.* 2019;20(10):2383. <https://doi.org/10.3390/ijms20102383> PMID: 31091705
6. Petta I, Lievens S, Libert C, Tavernier J, De Bosscher K. Modulation of protein-protein interactions for the development of novel therapeutics. *Mol Ther.* 2016;24(4):707–18. <https://doi.org/10.1038/mt.2015.214> PMID: 26675501
7. Yue Y, Li S, Cheng Y, Wang L, Hou T, Zhu Z, et al. Integration of molecular coarse-grained model into geometric representation learning framework for protein-protein complex property prediction. *Nat Commun.* 2024;15(1):9629. <https://doi.org/10.1038/s41467-024-53583-w> PMID: 39511202
8. Golub E, Subramanian RH, Esselborn J, Alberstein RG, Bailey JB, Chiong JA, et al. Constructing protein polyhedra via orthogonal chemical interactions. *Nature.* 2020;578(7793):172–6. <https://doi.org/10.1038/s41586-019-1928-2> PMID: 31969701
9. Rodrigues CHM, Pires DEV, Blundell TL, Ascher DB. Structural landscapes of PPI interfaces. *Brief Bioinform.* 2022;23(4):bbac165. <https://doi.org/10.1093/bib/bbac165> PMID: 35656714
10. Cheng F, Zhao J, Wang Y, Lu W, Liu Z, Zhou Y, et al. Comprehensive characterization of protein-protein interactions perturbed by disease mutations. *Nat Genet.* 2021;53(3):342–53. <https://doi.org/10.1038/s41588-020-00774-y> PMID: 33558758

11. Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, et al. A reference map of the human binary protein interactome. *Nature*. 2020;580(7803):402–8. <https://doi.org/10.1038/s41586-020-2188-x> PMID: [32296183](#)
12. Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci*. 2009;10(6):2763–88. <https://doi.org/10.3390/ijms10062763> PMID: [19582228](#)
13. Rajagopala SV, Sikorski P, Kumar A, Mosca R, Vlasblom J, Arnold R, et al. The binary protein-protein interaction landscape of *Escherichia coli*. *Nat Biotechnol*. 2014;32(3):285–90. <https://doi.org/10.1038/nbt.2831> PMID: [24561554](#)
14. Smyth MS, Martin JH. X ray crystallography. *Mol Pathol*. 2000;53(1):8–14. <https://doi.org/10.1136/mp.53.1.8> PMID: [10884915](#)
15. Lyu P, Iribarne M, Serjanov D, Zhai Y, Hoang T, Campbell LJ, et al. Common and divergent gene regulatory networks control injury-induced and developmental neurogenesis in zebrafish retina. *Nat Commun*. 2023;14(1):8477. <https://doi.org/10.1038/s41467-023-44142-w> PMID: [38123561](#)
16. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*. 2008;9(10):770–80. <https://doi.org/10.1038/nrm2503> PMID: [18797474](#)
17. Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*. 2019;35(12):2159–61. <https://doi.org/10.1093/bioinformatics/bty916> PMID: [30445495](#)
18. Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim J-S, et al. A novel signaling pathway impact analysis. *Bioinformatics*. 2009;25(1):75–82. <https://doi.org/10.1093/bioinformatics/btn577> PMID: [18990722](#)
19. Zhang Q, Wang L, Wang S, Cheng H, Xu L, Pei G, et al. Signaling pathways and targeted therapy for myocardial infarction. *Signal Transduct Target Ther*. 2022;7(1):78. <https://doi.org/10.1038/s41392-022-00925-z> PMID: [35273164](#)
20. Čapek D, Safroshkin M, Morales-Navarrete H, Toulany N, Arutyunov G, Kurzbach A, et al. EmbryoNet: using deep learning to link embryonic phenotypes to signaling pathways. *Nat Methods*. 2023;20(6):815–23. <https://doi.org/10.1038/s41592-023-01873-4> PMID: [37156842](#)
21. Kracht M, Müller-Ladner U, Schmitz ML. Mutual regulation of metabolic processes and proinflammatory NF-κB signaling. *J Allergy Clin Immunol*. 2020;146(4):694–705. <https://doi.org/10.1016/j.jaci.2020.07.027> PMID: [32771559](#)
22. Scarinci G, Ariens J-L, Angelidou G, Schmidt S, Glatter T, Paczia N, et al. Enhanced metabolic entanglement emerges during the evolution of an interkingdom microbial community. *Nat Commun*. 2024;15(1):7238. <https://doi.org/10.1038/s41467-024-51702-1> PMID: [39174531](#)
23. Li Y, Golding GB, Ilie L. DELPHI: accurate deep ensemble model for protein interaction sites prediction. *Bioinformatics*. 2021;37(7):896–904. <https://doi.org/10.1093/bioinformatics/btaa750> PMID: [32840562](#)
24. Wang F, Chu J, Shen L, Chang S. MESM: integrating multi-source data for high-accuracy protein-protein interactions prediction through multimodal language models. *BMC Biol*. 2025;23(1):253. <https://doi.org/10.1186/s12915-025-02356-y> PMID: [40784875](#)
25. Vella D, Marini S, Vitali F, Di Silvestre D, Mauri G, Bellazzi R. MTGO: PPI network analysis via topological and functional module identification. *Sci Rep*. 2018;8(1):5499. <https://doi.org/10.1038/s41598-018-23672-0> PMID: [29615773](#)
26. Grassmann G, Miotto M, Desantis F, Di Rienzo L, Tartaglia GG, Pastore A, et al. Computational approaches to predict protein-protein interactions in crowded cellular environments. *Chem Rev*. 2024;124(7):3932–77. <https://doi.org/10.1021/acs.chemrev.3c00550> PMID: [38535831](#)
27. Zhang J, Durham J, Qian Cong. Revolutionizing protein-protein interaction prediction with deep learning. *Curr Opin Struct Biol*. 2024;85:102775. <https://doi.org/10.1016/j.sbi.2024.102775> PMID: [38330793](#)
28. Mou M, Pan Z, Zhou Z, Zheng L, Zhang H, Shi S, et al. A transformer-based ensemble framework for the prediction of protein-protein interaction sites. *Research (Wash D C)*. 2023;6:0240. <https://doi.org/10.34133/research.0240> PMID: [37771850](#)
29. Hou Z, Yang Y, Ma Z, Wong K-C, Li X. Learning the protein language of proteome-wide protein-protein binding sites via explainable ensemble deep learning. *Commun Biol*. 2023;6(1):73. <https://doi.org/10.1038/s42003-023-04462-5> PMID: [36653447](#)
30. Gao Z, Jiang C, Zhang J, Jiang X, Li L, Zhao P, et al. Hierarchical graph learning for protein-protein interaction. *Nat Commun*. 2023;14(1):1093. <https://doi.org/10.1038/s41467-023-36736-1> PMID: [36841846](#)
31. Wu L, Tian Y, Huang Y, Li S, Lin H, Chawla NV, et al. MAPE-PPI: Towards Effective and Efficient Protein-Protein Interaction Prediction via Microenvironment-Aware Protein Embedding. In: *The Twelfth International Conference on Learning Representations*; 2024.
32. Lv G, Hu Z, Bi Y, Zhang S. Learning Unknown from Correlations: Graph Neural Network for Inter-novel-protein Interaction Prediction. In: *IJCAI-21*.
33. Kang Y, Elofsson A, Jiang Y, Huang W, Yu M, Li Z. AFTGAN: prediction of multi-type PPI based on attention free transformer and graph attention network. *Bioinformatics*. 2023;39(2):btad052. <https://doi.org/10.1093/bioinformatics/btad052> PMID: [36692145](#)
34. Tang T, Li T, Li W, Cao X, Liu Y, Zeng X. Anti-symmetric framework for balanced learning of protein-protein interactions. *Bioinformatics*. 2024;40(10):btad603. <https://doi.org/10.1093/bioinformatics/btae603> PMID: [39404784](#)
35. Zhao Z, Qian P, Yang X, Zeng Z, Guan C, Tam WL, et al. SemiGNN-PPI: Self-Ensembling Multi-Graph Neural Network for Efficient and Generalizable Protein-Protein Interaction Prediction. In: Elkind E, editor. *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23. International Joint Conferences on Artificial Intelligence Organization*; 2023. pp. 4984–92. <https://doi.org/10.24963/ijcai.2023/554>
36. Zeng X, Meng F-F, Wen M-L, Li S-J, Li Y. GNNGL-PPI: multi-category prediction of protein-protein interactions using graph neural networks based on global graphs and local subgraphs. *BMC Genomics*. 2024;25(1):406. <https://doi.org/10.1186/s12864-024-10299-x> PMID: [38724906](#)
37. RDKit Community. RDKit: Open-Source Cheminformatics Software. 2020. [cited 2020 Aug 7]. Available from: <https://www.rdkit.org>

38. Chen Y, Wang W, Liu J, Feng J, Gong X. Protein Interface Complementarity and Gene Duplication Improve Link Prediction of Protein-Protein Interaction Network. *Front Genet.* 2020;11:291. <https://doi.org/10.3389/fgene.2020.00291> PMID: [32300358](#)
39. Lv G, Hu Z, Bi Y, Zhang S. Learning unknown from correlations: graph neural network for inter-novel-protein interaction prediction. In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*. 2021. pp. 3577–684.
40. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryar F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 2023;51(D1):D638–46. <https://doi.org/10.1093/nar/gkac1000> PMID: [36370105](#)
41. Chen M, Ju CJ-T, Zhou G, Chen X, Zhang T, Chang K-W, et al. Multifaceted protein-protein interaction prediction based on Siamese residual RCNN. *Bioinformatics.* 2019;35(14):i305–14. <https://doi.org/10.1093/bioinformatics/btz328> PMID: [31510705](#)
42. Sun Y, Wang R, Luo Z, Tan L, Liu J, Li R, et al. ESM2_AMP: an interpretable framework for protein-protein interactions prediction and biological mechanism discovery. *Brief Bioinform.* 2025;26(4):bbaf434. <https://doi.org/10.1093/bib/bbaf434> PMID: [40874821](#)
43. Ko YS, Parkinson J, Liu C, Wang W. TUNA: an uncertainty-aware transformer model for sequence-based protein-protein interaction prediction. *Brief Bioinform.* 2024;25(5):bbae359. <https://doi.org/10.1093/bib/bbae359> PMID: [39051117](#)
44. Zhong W, He C, Xiao C, Liu Y, Qin X, Yu Z. Long-distance dependency combined multi-hop graph neural networks for protein-protein interactions prediction. *BMC Bioinformatics.* 2022;23(1):521. <https://doi.org/10.1186/s12859-022-05062-6> PMID: [36471248](#)
45. Ozenne B, Subtil F, Maucourt-Boulch D. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol.* 2015;68(8):855–9. <https://doi.org/10.1016/j.jclinepi.2015.02.010> PMID: [25881487](#)
46. Aros CJ, Pantoja CJ, Gomperts BN. Wnt signaling in lung development, regeneration, and disease progression. *Commun Biol.* 2021;4(1):601. <https://doi.org/10.1038/s42003-021-02118-w> PMID: [34017045](#)
47. Hayat R, Manzoor M, Hussain A. Wnt signaling pathway: a comprehensive review. *Cell Biol Int.* 2022;46(6):863–77. <https://doi.org/10.1002/cbin.11797> PMID: [35297539](#)
48. Bajaj R, Warner AN, Fradette JF, Gibbons DL. Dance of the Golgi: understanding Golgi dynamics in cancer metastasis. *Cells.* 2022;11(9):1484. <https://doi.org/10.3390/cells11091484> PMID: [35563790](#)
49. Del Toro N, Shrivastava A, Ragueneau E, Meldal B, Combe C, Barrera E, et al. The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.* 2022;50(D1):D648–53. <https://doi.org/10.1093/nar/gkab1006> PMID: [34761267](#)
50. Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with Cytoscape 3. *Curr Protoc Bioinformatics.* 2014;47:8.13.1–24. <https://doi.org/10.1002/0471250953.bi0813s47> PMID: [25199793](#)
51. Walsh CT, Garneau-Tsodikova S, Gatto Jr GJ. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angewandte Chemie Int Ed.* 2005;44(45):7342–72.
52. Cornish-Bowden A. *Fundamentals of enzyme kinetics*. John Wiley & Sons; 2013.