BMC Bioinformatics

RESEARCH Open Access

PPI-Graphomer: enhanced protein-protein affinity prediction using pretrained and graph transformer models

Jun Xie[1,2,3], Youli Zhang[2,3], Ziyang Wang[1,2,3], Xiaocheng Jin[2,3], Xiaoli Lu[4], Shengxiang Ge[2,3*] and Xiaoping Min[1,2,3*]

*Correspondence: sxge@xmu.edu.cn; mxp@xmu.edu.cn

[1] Institute of Artificial Intelligence, School of Informatic, Xiamen University, No. 422 Siming South Rd, Xiamen 361005, China
[2] National Institute of Diagnostics and Vaccine Development in Infectious Diseases, School of Public Health, Xiamen University, No. 422 Siming South Rd, Xiamen 361005, China
[3] State Key Laboratory of Vaccines for Infectious Diseases, Xiang An Biomedicine Laboratory, School of Public Health, Xiamen University, No. 422 Siming South Rd, Xiamen 361005, China
[4] Information and Networking Center, Xiamen University, No. 422 Siming South Rd, Xiamen 361005, China

Abstract

Protein-protein interactions (PPIs) refer to the phenomenon of protein binding through various types of bonds to execute biological functions. These interactions are critical for understanding biological mechanisms and drug research. Among these, the protein binding interface is a critical region involved in protein-protein interactions, particularly the hotspot residues on it that play a key role in protein interactions. Current deep learning methods trained on large-scale data can characterize proteins to a certain extent, but they often struggle to adequately capture information about protein binding interfaces. To address this limitation, we propose the PPI-Graphomer module, which integrates pretrained features from large-scale language models and inverse folding models. This approach enhances the characterization of protein binding interfaces by defining edge relationships and interface masks on the basis of molecular interaction information. Our model outperforms existing methods across multiple benchmark datasets and demonstrates strong generalization capabilities.

Keywords: Bind affinity prediction, ESM, Graph transformer

Introduction

Proteins are the primary executors of biological activities, and their functions are often exerted through interactions [1, 2]. Therefore, elucidating the mechanisms of protein-protein interactions is pivotal to protein studies. Binding affinity serves as a critical indicator of these interactions [3], as its magnitude signifies the potential for proteins to cooperate effectively. Understanding binding affinity also facilitates the identification of promising candidates in drug and inhibitor design [4–6], thereby enabling high-throughput screening and design processes.

The quantification of protein affinity is typically expressed via the equilibrium dissociation constant (Kd), which represents the ratio of the concentrations of dissociated and bound molecular states at equilibrium. A lower Kd value indicates tighter binding and stronger affinity. Due to the time-consuming nature and high material cost associated

with experimental determination of affinity [7], the accurate prediction of protein-protein affinity through computational methods is highly important. Among these methods, molecular dynamics simulations require substantial computational resources [8, 9] and are often too time-intensive for large-scale screening. Alternatively, empirical functions offer faster evaluations of affinity [10–12], as seen with knowledge-based statistical potential tools like DFIRE [13] and physics-based energy functions such as FoldX [14], or their combinations like RosettaDock [15]. However, these empirical functions are often constrained by the limitations of specific scenarios and are unable to update their discriminative capabilities based on data amplification. In contrast, machine learning and deep learning methods automatically extract more complex latent features through learning processes. Their performance can gradually improve as the scale of data expands, enabling more accurate predictions. Consequently, these approaches are receiving increasing attention. The application of machine learning and deep learning methodologies in the biological sciences has increasingly gained traction and achieved significant breakthroughs. Notable examples include the use of AlphaFold2 for precise protein structure prediction [16] and the application of RFdiffusion for de novo protein structure design [17]. These advances suggest that leveraging deep learning for the representation of proteins or protein interactions is feasible, thus heralding potential advancements for various downstream tasks in the protein research. Currently, several algorithms based on machine learning and deep learning have emerged in the domain of protein affinity prediction. The prediction of protein affinity via machine learning and deep learning methods can be categorized based on data sources into sequence-based and structure-based approaches [18]. Sequence-based methods often derive features from statistical analyses of amino acid frequency, conservation, and physicochemical properties within the sequence [19]. For instance, ISLAND [20] employs BLOSUM-based feature extraction kernels to capture sequence information such as optimal alignments, local alignments, and mismatches. Inspired by the remarkable capabilities of AlphaFold, some studies have harnessed structure predictions with AlphaFold2 to derive indicators such as interface predicted aligned error (iPAE) to represent affinity. Sequence-based methods generally provide more abundant data and are less constrained in application scenarios. However, they yield less detailed information compared to structure-based approaches and do not capture binding sites within complexes, thus rendering structure-based approaches as the prevailing methodology [21, 22].

For example, PRODIGY [23] extracts structural information through buried surface area and non-interacting surface (NIS) features, while CP-PIE [24] calculates overlap and solvent-accessible surface area as structural features. On the other hand, PPI-Affinity [25] synthesizes a structural feature set exceeding 20000 dimensions using a variety of feature extraction tools. These tasks frequently employ Support Vector Machines (SVM) as the regression model, as simpler models often perform better given the limited data volume. Consequently, advancements in these approach have primarily focused on innovations in feature extraction methods.

The efficacy of feature extraction is a pivotal determinant of the ultimate performance of predictive models. Due to the limited availability of protein-protein affinity data, it is generally challenging to derive features with high generalization capability. Pretrained models serve as powerful instruments for feature extraction [26]. Through

self-supervised training on large-scale datasets, these models can learn generalized patterns or regularities of proteins, thereby effectively mitigating the risk of overfitting in downstream tasks. In the field of protein research, several pretrained models, such as ESM2, ProtBert, and ESM-IF1, are available for this purpose. ESM2, developed by Meta, leverages a transformer architecture [27] and is pretrained on over 600 million protein sequences. This model effectively captures protein sequence information by constructing a protein language model [28] through a BERT-like [29] training approach, and it has demonstrated exceptional performance across various downstream tasks [30–32]. Additionally, ESM-IF1 use a GVP-GNN architecture and is pretrained on over 12 million AlphaFold2-predicted structural datasets, along with 16000 real data entries. This integration allows ESM-IF1 to achieve a sequence recovery rate of up to 51%, indicating a notably competitive outcome.

However, existing pre-trained models are inadequate for effectively capturing interaction information between complexes. Protein-protein interactions are often influenced by a few critical hotspot residues [33, 34], whereas current pre-trained models are predominantly trained on monomers, lacking insights into interaction interface information. Consequently, in this study, we introduce a protein-protein affinity prediction tool based on the PPI-Graphormer module. This approach employs ESM2 and ESM-IF1 to extract more generalizable sequence and structural features. Building upon these pretrained features, we examine intermolecular interaction forces to introduce edge relationships between interface amino acids and construct the Graphomer module, a graph transformer, to concurrently process sequence and graph information. This methodology enables the capture of potential interactions between complexes, facilitating precise affinity prediction.

Method

The primary network architecture utilized in this study is depicted in Fig. 1A, with the input and output for predictions defined in Eq. (1). This approach involves leveraging sequence and structural information extracted from pretrained models, combined with specific interface information, to predict affinity through deep learning algorithms. Sequence and structural features are extracted using ESM2 and ESM-IF1, respectively. For multi-chain complexes, a linker consisting of 25 glycine residues is employed to concatenate them into a single chain. The resulting feature representations are fed into the PPI-Graphomer layer, which is a biased transformer self-attention layer. The bias terms in this layer are derived from three types of encodings based on the protein interaction interface: amino acid pair type encoding, interaction force encoding, and interface masking. The method of introducing interaction information through bias terms is inspired by the Graphomer model [35], thus the PPI-Graphomer architecture can be viewed as a graph transformer network that integrates sequence and structural information.

Within the PPI-Graphomer, the interaction information of critical residues at the interface is effectively captured, leading to the derivation of interface representations. The original sequence and structural representations are then concatenated with the interface representation using a method analogous to skip connections [36], resulting in the final integrated feature (Eq. 1). A Multilayer Perceptron (MLP) is subsequently used to perform regression prediction of the affinity value. The loss is computed against the

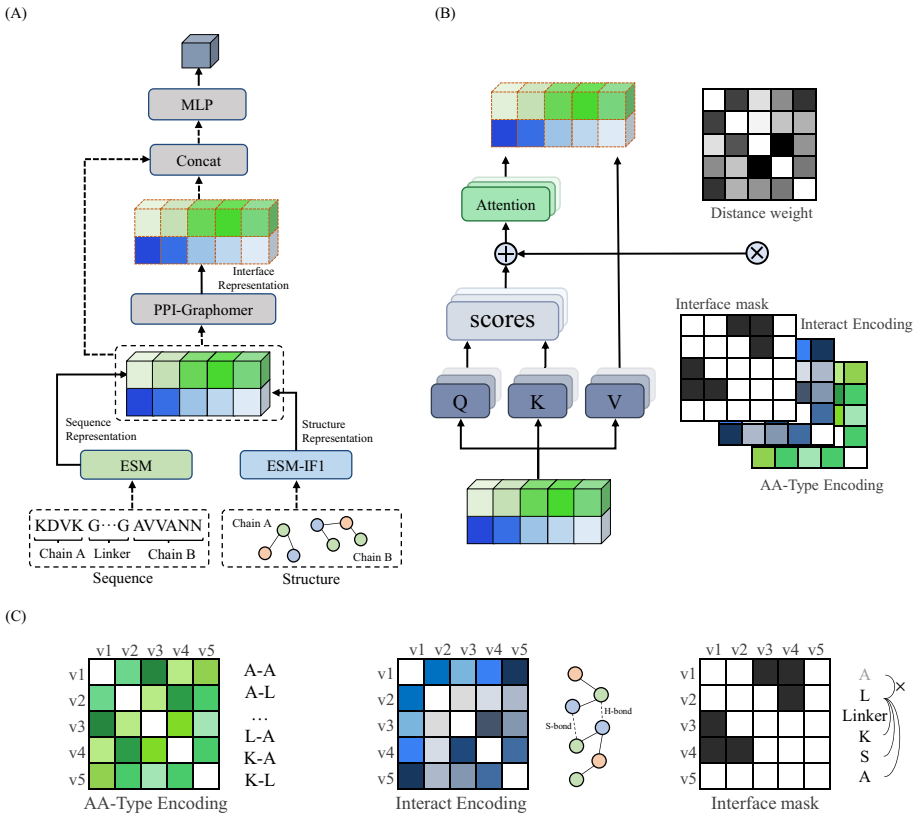

Fig. 1 The framework for protein-protein affinity prediction using pre-trained models and the PPI-Graphomer module. **A** The workflow of this study involves: utilizing ESM2 to extract sequence representations of protein complexes and ESM-IF1 to extract structural representations, which are subsequently concatenated. The concatenated representations are then processed through the PPI-Graphomer module to obtain interface representations. These interface representations are further concatenated with the original features. Ultimately, a MLP is employed to predict affinity values quantitatively. **B** The architecture of PPI-Graphomer incorporates bias terms based on three different encodings into the attention matrix, which are then multiplied by distance-based weight coefficients. **C** Three encodings based on amino acid structural information are utilized: encoding of amino acid pair types, encoding based on intermolecular interaction forces, and a masking matrix based on interface information

affinity labels (Eq. 2), and gradient backpropagation is executed. A detailed description of each component of the network will follow.

$$\Delta G_{pred} = f(x_{sequence}, x_{structure}, x_{\text{interface}}) \tag{1}$$

$$Loss = \frac{\sum_{i=1}^{n} \left| \Delta G_{label} - \Delta G_{pred} \right|}{n} \tag{2}$$

Sequence and structure representation

Protein features can generally be extracted from either sequence information or structural data. The amino acid sequence, as the primary structure of proteins, contains all the information necessary to form complex three-dimensional structures, which in turn dictate protein interactions and functions. Amino acid sequences are often described

as the language of proteins, because they share many characteristics with human languages. For instance, both have hierarchical structures [37], and protein folding rules bear some resemblance to grammatical syntax [38]. Consequently, applying natural language processing (NLP) techniques to protein sequences is a promising endeavor.

In this study, we utilize the ESM2 model to process protein sequences. By employing masked language modeling from NLP, ESM2 can capture evolutionary information and semantic features within protein sequences. We use only the output from the final encoder layer as the sequence represetation, without further training the network layers. Given the potential presence of chain breaks in the dataset due to missing amino acids, which can affect contextual understanding, we employ glycine residues to link broken chains and subsequently remove features at these positions. To ensure the correctness of multi-chain semantic information, we adopt a method similar to the ESMFold complex prediction approach, using 25 glycine residues as a linker to connect all chains into a single sequence. This method ensures consistency in the model and simplifies the processing flow. Post-feature extraction with ESM2, features at these linker positions are removed.

As complex macromolecular structures, proteins encode rich interaction information in the spatial arrangement of their amino acids. Graph neural networks (GNNs) are capable of capturing such structural information, yet they must address issues of rotational and translational invariance. GVP-GNN achieves precise capture of geometric information through geometric vector perceptrons. Based on GVP-GNN, ESM-IF1 constructs a structure-to-sequence predictor and pre-trained using span masking to extract structural information. We utilize the output from its encoder as the structural representation for downstream tasks.

To ensure alignment of features extracted by different tools, we exclude non-standard amino acids and those lacking C, N, and Ca backbone atoms, since ESM2 struggles with non-standard residues and ESM-IF1 requires these three backbone atoms for feature extraction. Ultimately, the sequence and structural representations are concatenated as input for subsequent processes.

PPI-Graphomer architecture

In protein-protein interactions, certain hotspot residues at the interface play a pivotal role in maintaining the tight binding between two proteins. Therefore, in the context of affinity prediction, it is crucial to focus on the amino acid characteristics at these interfaces. Existing pretrained models typically treat multi-chain outputs as single-chain inputs, encoding only general interaction information and lack the ability to adequately capture inter-chain interface details. Thus, we introduce the PPI-Graphormer architecture to specifically learn the interactions between different chains. The network architecture is based on Graphormer, a model developed by Microsoft that integrates graph structures into the transformer model. It primarily employs three types of encodings to incorporate edge information of the graph network.

We conceptualize amino acid interactions as edge information within a graph network. Based on this conceptualization, we design two encoding schemes derived from structure information: the $b(v_i, v_j)$ based on amino acid types and the $c(v_i, v_j)$ on intermolecular interaction forces. Additionally, the distance between amino acid pairs is used

as a weighting factor(D_{ij}). The product of these factors is incorporated as a bias term in the attention matrix of the self-attention layer, thereby enhancing the model's focus on interface information (Eq. 3).

$$A_{ij} = \frac{\left(h_i w_Q\right)\left(h_j w_k\right)^T}{\sqrt{d}} + D_{ij} * \left(b\left(v_i, v_j\right) + c\left(v_i, v_j\right)\right) \tag{3}$$

Considering that specific amino acids engage in characteristic interactions-such as hydrophobic amino acids aggregating to form a hydrophobic core, or cysteines forming disulfide bonds-we encode all types of amino acid pairs using one-hot or embedding representations to capture these interactions. The encoding of amino acid pair types is inspired by the spatial encoding used in Graphormer, originally mapping an integer, which indicates the shortest path length, to a high-dimensional embedding. While this integer type might reflect interaction strength, the embedding approach treats it predominantly as a categorical variable. Therefore, we employ amino acid pair types directly as a measure of this categorical variable. Different amino acid pairs may imply varying interaction intensities, which also serves as an approximate estimate of interaction strength.

We constructed a matrix with dimensions corresponding to the amino acid length squared, where each element represents a type of amino acid pair. Without considering the directionality of amino acid pairs, there are a total of 210 types of amino acid pairs $AAtype\left(v_i, v_j\right)$. A learnable embedding method is then used to map this matrix into the dimensional space corresponding to the number of attention heads, facilitating the learning of high-dimensional features of different amino acid pairs and enabling direct addition to the attention matrix (Eq. 4).

$$b\left(v_i, v_j\right) = Embedding(AAtype\left(v_i, v_j\right)) \tag{4}$$

The interaction force encoding is derived from edge encoding, where it learns weights for edge features along the shortest connecting path to determine the interaction strength between nodes. This process involves a detailed consideration of all information along the shortest path to further quantify interactions. Similarly, we meticulously evaluate potential interactions within amino acid pairs, sourced from various intermolecular interaction forces. Beyond amino acid types, the positional relationships between amino acids also influence potential intermolecular interactions. Therefore, we incorporate amino acid distances to tally the number of different intermolecular interaction forces, abstracting these as edge information for the amino acids. To simplify the data processing, we focus only on possible hydrogen bonds, halogen bonds, disulfide bonds, salt bridges, and π-π stacking.

By considering different intermolecular interaction forces, we construct a matrix with dimensions corresponding to amino acid length by amino acid length by the number of interaction force types x_{ij}^n. Each element in this matrix signifies the number of corresponding atomic interaction force types between that amino acid pair. Parameterizing intermolecular interactions solely through numerical quantities is relatively simplistic. In the future, we plan to integrate more bioinformatics tools to explore and assign distinct weights to different intermolecular interaction forces, which is an area we aim to

further investigate. Subsequently, a linear layer is employed to learn weights w_e for different atomic interaction force types, mapping them into the dimensional space corresponding to the number of attention heads, facilitating addition to the attention matrix (Eq. 5).

$$c(v_i, v_j) = \sum_{n=1}^{N} x_{ij}^{n} (w_e)^T \tag{5}$$

The resulting two encoding matrices serve as bias terms in the attention matrix, representing enhanced focus on specific amino acids. To ensure that this additional focus has a realistic significance, we multiply the bias term by a distance coefficient. This distance coefficient is based on a threshold of 7 Å, where the weight decreases as the distance approaches 7 Å and increases as the distance approaches 0, suggesting a stronger potential interaction. Interactions beyond 7 Å are directly excluded by setting the attention to zero through a mask matrix. Additionally, to consider only interface interactions between different chains, all amino acids within the same chain are masked. Thus, only attention between amino acids from different chains within the 7 Å range is ultimately considered (Eq. 6).

$$\text{Mask}(i,j) = \begin{cases} 1 \text{ if } D_{ij} \leq 7\text{Å and Chain}_i \neq \text{Chain}_j \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

By integrating the aforementioned two encoding and employing masking on non-interface amino acids, we embed structural information into the sequence representation. This approach underscores the critical role of hot spot residues at the interface in protein-protein interactions, resulting in an interface representation that is utilized for downstream tasks.

Representation concat and affinity predict

The sequence and structural representations extracted via ESM2 and ESM-IF1 are processed through the PPI-Graphomer to be transformed into interface representations. However, this representation alone cannot serve as the sole source of information for affinity prediction, as the computation of interface information results in the loss of the majority of the original sequence or structural data. Since the dimension of the interface representation matches that of the original input, we employ an approach akin to skip connections by directly concatenating the input and output of the PPI-Graphomer module. This enables the simultaneous consideration of both the complete and interfaced information. Subsequently, a MLP is applied for affinity prediction.

The average feature across all amino acids is considered as the protein feature. Subsequently, a MLP is employed to learn from this feature, mapping it to a one-dimensional representation. The loss is computed relative to the affinity label, followed by gradient backpropagation.

Experiment

Dataset

The dataset utilized in this study primarily originates from PDBbind [39], a comprehensive database of protein complexes that includes PPI structures and experimentally determined affinities. The most recent version released in 2020 contains 2,852 samples, each consisting of structural coordinate information recorded in PDB files and experimentally measured affinity values, expressed as Kd, Ki, or IC50. For ease of comparison, these affinity labels have been uniformly converted to Gibbs free energy. To facilitate comparative analysis, the test set adopted in this study aligns with that used in reference [25], derived from both the "structure-based benchmark database for protein-protein binding affinity" [40](Test set 1) and a benchmark set extracted from PDBbind(Test set 2). To further evaluate the performance, we also conducted tests on the new version of the test set 1, namely "Affinity Benchmark Version 2" [41]. The test set 1 comprises 79 samples, and the test set 2 includes 90 samples; after removing samples with residues sequence exceeding a length of 2000, 75 and 87 samples remain respectively. Since the reduction in the number of samples is minimal, we believe that removing these samples will not significantly affect the final results. Additionally, Complexes with residues sequence exceeding a length of 2000 often contain multiple subunits that may not play a critical role in interface binding in all analyses. To facilitate a unified comparison of the distribution characteristics across different datasets, we constructed density plots Fig. 2. It can be observed that they exhibit similar peak values and some overlapping regions, indicating that their distributions may be similar. The frequency histograms for each dataset can be found in the supplementary materials.

The utilization of sequence models necessitates the removal of excessively long sequences from the dataset to prevent a significant increase in model complexity; thus,

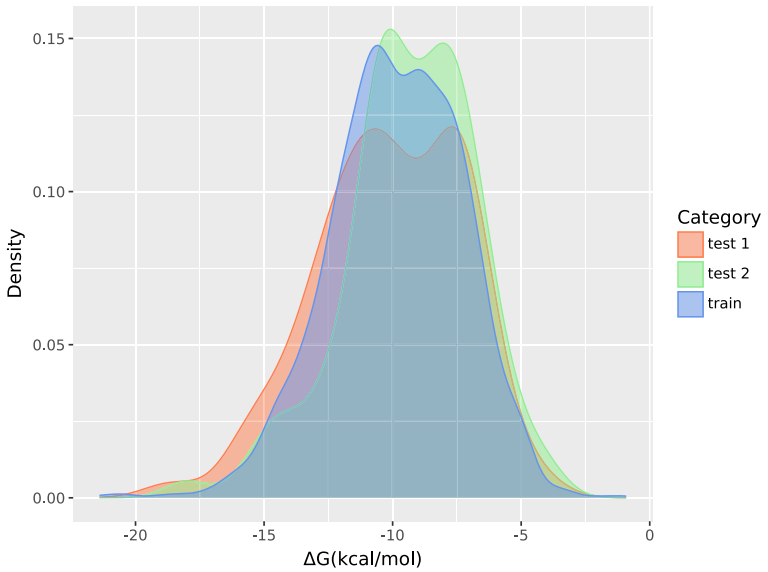

Fig. 2 The density plots of the ΔG for the training set and two test sets. The x-axis represents the label values, while the y-axis denotes the probability distribution of the labels. The choice of a density plot over a histogram is attributed to the former's ability to provide a smoother distribution estimation and facilitate the observation of distributions from datasets of varying sizes within the same figure

residues sequence exceeding a length of 2000 were excluded. Additionally, to mitigate the risk of data leakage, deduplication of the existing data was conducted. This deduplication process was implemented in two parts: first, targeting the aforementioned test set, and second, the dataset intended for cross-validation. To prevent overlap between the training and test sets, BLAST [42] was employed to remove samples from the training set that exhibited high similarity to those in the test set. The criteria for deduplication required the similarity to exceed 0.65 and the overlapping sequence length to constitute more than 80% of the total sequence length. Given that complexes may exhibit high similarity on one side and not on the other, leading to small length ratios but high similarity in repeated sequences, we considered such complexes to be valid distinct samples. Therefore, by setting a length ratio threshold, these samples were retained in the training set, resulting in a final training set comprising 2376 samples. Furthermore, to accurately evaluate model performance, a dataset for 5-fold cross-validation was constructed, derived from the collective pool of all samples. Utilizing CD-HIT [43] for deduplication, with a similarity threshold set at 0.75, the 5-fold cross-validation dataset achieved a size of 2085 samples.

Pre-train model

In this study, the ESM2 model [44] was employed to extract sequence features, while ESM-IF1 [45] was used to extract structural features. Biopython was utilized to analyze intermolecular interaction forces. ESM2 comprises multiple pretrained weights of varying scales, with encoder layer counts from 6 to 48, covering parameter sizes from 8 M to 15B, and producing embeddings with dimensions from 320 to 5120. To balance training efficiency and performance, we selected the model with 650 M parameters, using the output from the 33rd encoder layer as the final feature representation, yielding a feature dimension of 1280. To reduce the parameter count of model and prevent overfitting, a linear layer was employed to map the 1280-dimensional features to 64 dimensions before concatenating them with other features.

In ESM-IF1, the encoder output is utilized as the feature representation with a dimension of 512. Similar to ESM2, a linear layer mapped this to a lower dimension. Given the lower output dimensionality of ESM-IF1, it was mapped to 32 dimensions. Then, the outputs from these two pretrained models were standardized to ensure comparability, enhancing model stability and training speed.

Model details

Based on the outputs of the pretrained models and the PPI-Graphormer architecture, we constructed a two-layer PPI-Graphormer module that incorporates attention biases. This module features an embedding dimension of 96, with 8 attention heads and Q, K, V (query, key, value) dimension of 32. The Adam optimizer was employed, with the learning rate initialized at 8e-4 and scheduled to decay by a factor of 0.95 each epoch. The model was trained for a total of 20 epochs on a single A40 GPU, and any GPU with 4GB of memory can be used for inference.

Evaluation metric

To evaluate the model's performance, we employed Mean Absolute Error (MAE) and the Pearson Correlation Coefficient (PCC) as primary metrics. These metrics are defined as follows: MAE quantifies the average magnitude of errors between predicted values and actual binding affinities across N samples, as calculated by the following formula:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{7}$$

The PCC measures the linear correlation between predicted values and actual binding affinities, providing insight into the accuracy and directionality of predictions. The calculation method is given by the following formula:

$$PCC = \frac{\sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n} (x_i - \bar{Y})^2}} \tag{8}$$

To facilitate comparison with previous studies and because the Kd values provided in the dataset exhibit an exponential relationship that complicates loss calculation, we convert the equilibrium dissociation constants into Gibbs free energy using the formula. In this conversion, R denotes the universal gas constant ($R = 8.314\ J{\cdot}mol^{-1} \cdot K^{-1}$), and T is set to room temperature at 298 K (25 °C).

$$\Delta G = RT \ln Kd \tag{9}$$

Result

5-fold evaluation

To comprehensively evaluate the model's performance, we conducted cross-validation on the previously mentioned dataset, which consists of 2085 samples and was divided using a 5-fold partitioning scheme. This approach segments the dataset into five equal parts, consecutively using one part as the validation set while the remaining portions are utilized for training. The process is repeated five times, ensuring each subset is used once for validation. During each training iteration, the PCC and MAE on the validation set were recorded. The average of these five iterations was calculated to yield the final result. Our results from the 5-fold cross-validation were a PCC of 0.581 and a MAE of 1.63.

Comparison with other methods on the benchmark set

To rigorously evaluate the performance of our model, we conducted comparative validation against other methods on the two aforementioned test sets. The results of this comparison are presented in Table 1 and Fig. 3. In the comparative analysis, the highest-performing metrics are highlighted in bold. The first test set comprises 75 samples with binding energy values ranging from 4.3 to 18.6. On this dataset, our model achieved a performance of PCC = 0.641 and MAE = 1.64, ranking second only to PRODIGY. Meanwhile, in [46], additional methods' performances on this dataset have been summarized.

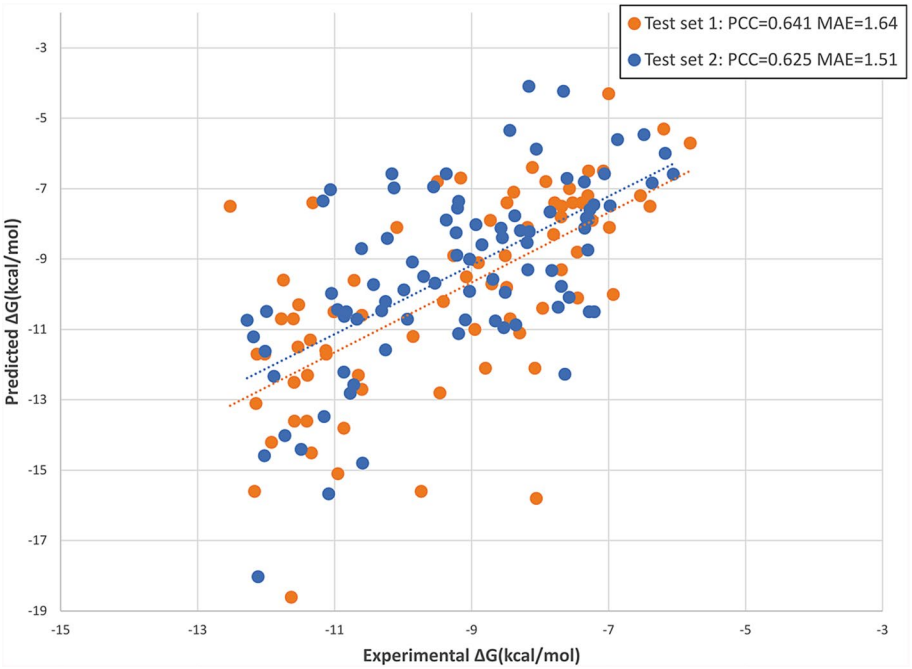

Fig. 3 The Scatter plot of predicted vs experimental binding affinities. The model performance was validated on two separate test datasets. The first test set comprised 75 samples, yielding a PCC of 0.641 and a MAE of 1.64. The second test set included 87 samples, achieving a PCC of 0.625 and an MAE of 1.51

Table 1 Comparison of model performances on the 2 test set and their combination

Method	Test set 1[1]		Test set 2[2]		Combined set[3]	
	PCC↑	MAE↓	PCC↑	MAE↓	PCC↑	MAE↓
PRODIGY	**0.735**	1.43	0.306	2.51	0.446	2.00
DFIRE	0.602	4.64	0.095	25.37	0.032	15.48
CP_PIE	−0.517	**0.80**	0.095	7.59	0.180	8.18
ISLAND	0.378	2.10	0.276	2.18	0.338	2.13
PPI-Affinity	0.616	1.82	0.495	1.80	0.559	1.80
ours	0.641	1.64	**0.625**	**1.51**	**0.633**	**1.57**

[1]The dataset derived from both the "structure-based benchmark database for protein-protein binding affinity" [11], comprises 75 samples

[2]The dataset extracted from PDBbind, comprises 87 samples

[3]The combined dataset from the aforementioned two datasets contains 162 samples

We calculated the PCC between the predicted results of these methods and the labels. Since different methods discard a small number of samples during data processing, we constructed their subsets for comparison. The results of this comparison are presented in Supplementary Table S1. We further validated our approach on the updated version of this dataset, which contains 188 samples after deduplication. The test results on these samples yielded a PCC of 0.626 and a MAE of 1.60. These results indicate that the performance did not exhibit a significant decline on a larger sample size.

We further validated our model on a larger test set, consisting of 87 samples introduced by the PPI-Affinity, with binding energy ranging from [4.03, 18.0]. Compared to the first dataset, this set exhibits a broader distribution of binding energies, which poses

Table 2 Performance comparison with other methods on binary complex dataset

Method	Test set 1	Test set 2
	PCC↑	PCC↑
Rosetta-InterfaceAnalyze[1]	0.482	0.054
Foldx-AnalyseComplex[2]	0.076	−0.159
ColabFold-iPAE[3]	0.588	0.156
ours	**0.708**	**0.633**

[1]The dG_separated score generated by RosettaFold's InterfaceAnalyzer to indicate binding affinity. Prior to this, the PDB files were processed using relax and jd_score2

[2]The Interaction Energy score generated by FoldX's AnalyseComplex module to indicate binding affinity

[3]The iPAE yielded by ColabFold to indicate binding affinity

Table 3 The ablation study results obtained by attempting the removal of various components

Method	Test set 1		Test set 2		Combined set	
	PCC↑	MAE↓	PCC↑	MAE↓	PCC↑	MAE↓
Standard	0.641	1.64	0.625	1.51	0.633	1.57
Without PPI-Graphomer	0.624	1.70	0.586	1.64	0.601	1.71
Without esm2 feature	0.545	1.83	0.521	1.69	0.511	1.91
Without esm-if1 feature	0.608	1.70	0.603	1.63	0.605	1.66

challenges for accurate prediction by many methods. Our model, however, attained a PCC of 0.625 and a MAE of 1.51 on this test set, achieving the best performance. When evaluated on the combined dataset of both test sets, our model achieved a PCC of 0.633 and a MAE of 1.57, also securing the top ranking.

To further verify the superiority of our model over traditional methods, we compared it against several empirical function-based methods on the test sets. The results are presented in Table 2. Bold values denote the best performance across methods. Since the outputs of these functions may not strictly conform to the definition of free energy, only the PCC was calculated, and MAE was omitted. We utilized Rosetta and FoldX for predictions. Due to their InterfaceAnalyzer and AnalyseComplex capabilities being limited to binary complexes, we selected subsets containing only binary complexes from the test sets for evaluation. The sizes of the two subsets were 23 and 78, respectively.

On the binary complex subsets, the PCC achieved by our model on the two test sets were 0.708 and 0.633, respectively. In contrast, the results from Rosetta's InterfaceAnalyzer were 0.482 and 0.054, while FoldX yielded scores of −0.076 and −0.159. Additionally, we employed a lightweight AlphaFold variant, ColabFold, to assess the two subsets. By extracting sequences from the PDB files, predicting structures, and using the iPAE as an affinity estimate, the obtained results were 0.588 and 0.156, respectively.

Ablation experiment

To validate the effectiveness of the model framework, we conducted ablation studies. The primary aim of this investigation was to analyze the significance of various features within the model and assess the impact of the PPI-Graphormer module on overall model performance, thereby elucidating their contributions to predicting protein-protein binding affinity. Our ablation experiments included the following components: removal of

features extracted by ESM2, removal of features extracted by ESM-IF1, and exclusion of the PPI-Graphormer module. To ensure consistency, the same training parameters, datasets, and metrics were employed throughout the experiments.

As shown in the ablation study results (Table 3), the sequence features extracted by ESM2 play a crucial role, with a significant decline in performance observed when these features are removed. In contrast, the structural features extracted by ESM-IF1 offer only limited performance improvement. This potential explanation is that ESM2 is trained on a larger data scale and possesses higher feature dimensionality, which implicitly encodes sufficient structural information. In contrast, ESM-IF1 is trained on a smaller sample size and similarly lacks adequate complex data, limiting its ability to supplement the patterns of interface interactions. Furthermore, the inclusion of the PPI-Graphormer module provides additional performance benefits. By employing a skip connection-like approach, the PPI-Graphormer module underscores the interaction information of interface residues, further enhancing model performance.

Conclusion

Protein-protein affinity prediction holds a crucial position in the study of protein interactions. However, the limited availability of datasets in this domain significantly constrains the application of deep learning. As large-scale models continue to advance, the underlying principles of protein interactions are expected to be captured through self-supervised training on extensive datasets, addressing various downstream tasks based on protein interactions. In this study, we employed large models, ESM2 and ESM-IF1, to extract sequence and structural features, effectively capturing evolutionary-scale and spatial information of proteins. To address the inadequate capability of large models in capturing interface interactions, we proposed the PPI-Graphormer module based on chemical bonding. Utilizing a training set derived from PDBbind, our model demonstrated robust generalization performance when compared with other established methods.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-025-06123-2.

Supplementary material 1

Author contributions

X.J. and M.X. provided the main idea and checked the manuscript, X.J. built and trained models, X.J. conducted the main experimental verification and wrote the main manuscript text, W.Z. provided experimental ideas and verified the content. Z.Y., J.X. and L.X. checked the article format and language. M.X. and G.S. supervised each process, checked and modified the manuscript. All authors reviewed the manuscript.

Funding

This research was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0501001), National Natural Science Foundation of China (62272399), Major Science and Technology Project of Fujian Provincial Health Commission (2021ZD01006) and the Fundamental Research Funds for the Central Universities (20720220006).

Availability of data and materials

The datasets presented in this study can be found in online repositories. The PDBbind dataset was downloaded from PDBbind. The code, data and model parameters used in this study were open-sourced from Code of PPI-Graphomer.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no Conflict of interest.

Received: 14 November 2024 Accepted: 28 March 2025
Published online: 29 April 2025

References

1. Lu H, Zhou Q, He J, Jiang Z, Peng C, Tong R, Shi J. Recent advances in the development of protein-protein interactions modulators: mechanisms and clinical trials. Signal Transduct Target Ther. 2020;5(1):213.
2. Peng X, Wang J, Peng W, Wu F-X, Pan Y. Protein-protein interactions: detection, reliability assessment and applications. Brief Bioinform. 2017;18(5):798–819.
3. Aloy P, Russell RB. Structural systems biology: modelling protein interactions. Nat Rev Mol Cell Biol. 2006;7(3):188–97.
4. Kaczor AA, Bartuzi D, Stępniewski TM, Matosiuk D, Selent J. Protein-protein docking in drug design and discovery. Comput Drug Discov Design. 2018;2018:285–305.
5. Jubb H, Higueruelo AP, Winter A, Blundell TL. Structural biology and drug discovery for protein-protein interactions. Trends Pharmacol Sci. 2012;33(5):241–8.
6. Scott DE, Bayly AR, Abell C, Skidmore J. Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. Nat Rev Drug Discov. 2016;15(8):533–50.
7. Zhou M, Li Q, Wang R. Current experimental methods for characterizing protein-protein interactions. ChemMedChem. 2016;11(8):738–56.
8. De Paris R, Quevedo CV, Ruiz DD, Souza O, Barros RC. Clustering molecular dynamics trajectories for optimizing docking experiments. Comput Intell Neurosci. 2015;2015(1): 916240.
9. Flower DR, Phadwal K, Macdonald IK, Coveney PV, Davies MN, Wan S. T-cell epitope prediction and immune complex simulation using molecular dynamics: state of the art and persisting challenges. Immunome Res. 2010;6:1–18.
10. Zhang C, Liu S, Zhu Q, Zhou Y. A knowledge-based energy function for protein- ligand, protein- protein, and protein- DNA complexes. J Med Chem. 2005;48(7):2325–35.
11. Kastritis PL, Bonvin AM. Are scoring functions in protein- protein docking ready to predict interactomes? clues from a novel binding affinity benchmark. J Proteome Res. 2010;9(5):2216–25.
12. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci. 2002;99(22):14116–21.
13. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci. 2002;11(11):2714–26.
14. Delgado J, Radusky LG, Cianferoni D, Serrano L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. Bioinformatics. 2019;35(20):4168–9.
15. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol. 2003;331(1):281–99.
16. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with alphafold. Nature. 2021;596(7873):583–9.
17. Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, Ahern W, Borst AJ, Ragotte RJ, Milles LF, et al. De novo design of protein structure and function with rfdiffusion. Nature. 2023;620(7976):1089–100.
18. Guo Z, Yamaguchi R. Machine learning methods for protein-protein binding affinity prediction in protein design. Front Bioinf. 2022;2:1065703.
19. Yugandhar K, Gromiha MM. Protein-protein binding affinity prediction from amino acid sequence. Bioinformatics. 2014;30(24):3583–9.
20. Abbasi WA, Yaseen A, Hassan FU, Andleeb S, Minhas FUAA. Island: in-silico proteins binding affinity prediction using sequence information. BioData Mining. 2020;13:1–13.
21. Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein–protein complexes. Elife. 2015;4:07454.
22. Liu X, Luo Y, Li P, Song S, Peng J. Deep geometric representations for modeling effects of mutations on protein-protein binding affinity. PLoS Comput Biol. 2021;17(8):1009284.
23. Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. Prodigy: a web server for predicting the binding affinity of protein-protein complexes. Bioinformatics. 2016;32(23):3676–8.
24. Ravikant D, Elber R. Pie-efficient filters and coarse grained potentials for unbound protein-protein docking. Proteins: Struct, Funct, Bioinf. 2010;78(2):400–19.
25. Romero-Molina S, Ruiz-Blanco YB, Mieres-Perez J, Harms M, Münch J, Ehrmann M, Sanchez-Garcia E. PPI-affinity: a web tool for the prediction and optimization of protein-peptide and protein-protein binding affinity. J Proteome Res. 2022;21(8):1829–41.
26. Qiu X, Sun T, Xu Y, Shao Y, Dai N, Huang X. Pre-trained models for natural language processing: a survey. Sci China Technol Sci. 2020;63(10):1872–97.

27. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. (2017) Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17, pp. 6000–6010. Curran Associates Inc., Red Hook, NY, USA
28. Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, Guo D, Ott M, Zitnick CL, Ma J, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci. 2021;118(15):2016239118.
29. Devlin J. Bert: Pre-training of deep bidirectional transformers for language understanding;2018. arXiv preprint arXiv:1810.04805
30. Verkuil R, Kabeli O, Du Y, Wicky BI, Milles LF, Dauparas J, Baker D, Ovchinnikov S, Sercu T, Rives A. Language models generalize beyond natural proteins. BioRxiv. 2022;22:2022.
31. Hie B, Candido S, Lin Z, Kabeli O, Rao R, Smetanin N, Sercu T, Rives A. A high-level programming language for generative protein design. BioRxiv. 2022;12:2022.
32. Rao R, Meier J, Sercu T, Ovchinnikov S, Rives A. Transformer protein language models are unsupervised structure learners. Biorxiv. 2020;22:2020.
33. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. Science. 1995;267(5196):383–6.
34. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces11edited by j. wells. J Mol Biol. 1998;280(1):1–9.
35. Ying C, Cai T, Luo S, Zheng S, Ke G, He D, Shen Y, Liu T-Y. Do transformers really perform badly for graph representation? Adv Neural Inf Process Syst. 2021;34:28877–88.
36. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 770–778
37. Ferruz N, Höcker B. Controllable protein design with language models. Nat Mach Intell. 2022;4(6):521–32.
38. Hu B, Xia J, Zheng J, Tan C, Huang Y, Xu Y, Li SZ. Protein language models and structure prediction: connection and progression;2022. arXiv preprint arXiv:2211.16742
39. Wang R, Fang X, Lu Y, Wang S. The pdbbind database: collection of binding affinities for protein- ligand complexes with known three-dimensional structures. J Med Chem. 2004;47(12):2977–80.
40. Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, Janin J. A structure-based benchmark for protein-protein binding affinity. Protein Sci. 2011;20(3):482–91.
41. Vreven T, Moal IH, Vangone A, Pierce BG, Kastritis PL, Torchala M, Chaleil R, Jiménez-García B, Bates PA, Fernandez-Recio J, et al. Updates to the integrated protein-protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2. J Mol Biol. 2015;427(19):3031–41.
42. Ye J, McGinnis S, Madden TL. Blast: improvements for better sequence analysis. Nucleic acids research. 2006;34(suppl_2):6–9.
43. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658–9.
44. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Smetanin N, Verkuil R, Kabeli O, Shmueli Y, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science. 2023;379(6637):1123–30.
45. Hsu C, Verkuil R, Liu J, Lin Z, Hie B, Sercu T, Lerer A, Rives A (2022) Learning inverse folding from millions of predicted structures. In: International Conference on Machine Learning, PMLR, pp 8946–8970
46. Moal IH, Jiménez-García B, Fernández-Recio J. Ccharppi web server: computational characterization of protein-protein interactions from structure. Bioinformatics. 2015;31(1):123–5.

Publisher's Note