RESEARCH

Open Access

An interpretable deep geometric learning model to predict the effects of mutations on protein–protein interactions using large-scale protein language model

Caiya Zhang¹, Yan Sun^{1,2,3} and Pingzhao Hu^{1,2,3,4,5,6*}

Abstract

Protein-protein interactions (PPIs) are central to the mechanisms of signaling pathways and immune responses, which can help us understand disease etiology. Therefore, there is a significant need for efficient and rapid automated approaches to predict changes in PPIs. In recent years, there has been a significant increase in applying deep learning techniques to predict changes in binding affinity between the original protein complex and its mutant variants. Particularly, the adoption of graph neural networks (GNNs) has gained prominence for their ability to learn representations of protein-protein complexes. However, the conventional GNNs have mainly concentrated on capturing local features, often disregarding the interactions among distant elements that hold potential important information. In this study, we have developed a transformer-based graph neural network to extract features of the mutant segment from the three-dimensional structure of protein-protein complexes. By embracing both local and global features, the approach ensures a more comprehensive understanding of the intricate relationships, thus promising more accurate predictions of binding affinity changes. To enhance the representation capability of protein features, we incorporate a large-scale pre-trained protein language model into our approach and employ the global protein feature it provides. The proposed model is shown to be able to predict the mutation changes in binding affinity with a root mean square error of 1.10 and a Pearson correlation coefficient of near 0.71, as demonstrated by performance on test and validation cases. Our experiments on all five datasets, including both single mutant and multiple mutant cases, demonstrate that our model outperforms four state-of-the-art baseline methods, and the efficacy was subjected to comprehensive experimental evaluation. Our study introduces a transformer-based graph neural network approach to accurately predict changes in protein-protein interactions (PPIs). By integrating local and global features and leveraging pretrained protein language models, our model outperforms state-of-the-art methods across diverse datasets. The results of this study can provide new views for studying immune responses and disease etiology related to protein mutations. Furthermore, this approach may contribute to other biological or biochemical studies related to PPIs.

Scientific contribution Our scientific contribution lies in the development of a novel transformer-based graph neural network tailored to predict changes in protein–protein interactions (PPIs) with excellent accuracy. By seamlessly integrating both local and global features extracted from the three-dimensional structure of protein–protein complexes,

*Correspondence: Pingzhao Hu phu49@uwo.ca Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.





and leveraging the rich representations provided by pretrained protein language models, our approach surpasses existing methods across diverse datasets. Our findings may offer novel insights for the understanding of complex disease etiology associated with protein mutations. The novel tool can be applicable to various biological and biochemical investigations involving protein mutations.

Keywords Graph neural network (GNN), Graph transformer, Protein–protein interaction (PPI), Binding affinity, Mutation changes

Introduction

Protein–protein interactions (PPIs) are crucial for many fundamental biological processes. Among them, PPIs induced by mutations play a critical role in understanding the mechanisms of some signaling pathways, immune responses, or the structural integrity of cellular components [1]. For instance, antibodies, which are a major component of human's immune system, interact with specific target antigens to trigger an immune response. There are studies suggesting that research on these interactions may aid in understanding how specific mutations will affect protein stability and help to study potential genetic susceptibility [1].

Binding affinity, or free energy of binding, is a commonly recognized thermodynamic criterion for measuring PPIs. Since wet-lab experiments are labor-intensive and time-consuming, emphasis has been placed on rapid and accurate automated methods. In particular, some machine learning methods, such as gradient boosting trees and support vector machines, offer the possibility to directly establish the relationship between mutations and binding affinity changes. Moreover, with the increasing complexity of data, there has been a growing interest in deep learning methods. So far, these automated methods fall into two main categories: sequence-based methods and structure-based methods [2]. Sequence-based methods, such as Stacked Auto-encoder, proposed by Sun et al. have been shown to achieve good prediction performance [3]. Nevertheless, structural information of PPIs, such as their locations and adjacent nodes, has also been shown to be important in PPI prediction [4]. As a result, deep learning frameworks incorporating structural features of proteins for prediction have been increasingly developed. For example, TopGBT [5] employs topologybased features to depict the protein complexes, whereas, this type of features are not initially intended to represent the interatomic interactions, thereby constraining their predictive capacity for the alterations in binding affinity arising from mutations or the detection of nuanced alterations in conformations due to their topological abstraction. The study of mCSM-PPI2 [6] integrates the well-established mutation Cut-off Scanning Matrix (mCSM) graph-based signatures framework with seven features including graph-based structure, evolutionary information, and non-covalent interaction network analysis to predict the effect of mutations on PPIs. The famous MutaBind2 [7] approach incorporates seven features, including the interaction between proteins, the evolutionary conservation of proteins, and the thermodynamic stability of protein complexes, etc.

Despite the growing interest in using machine learning to understand the structure of proteins, this task presents unique challenges that differ from other types of data since protein structures cannot be directly reduced to simple numerical or pixel representations. Unlike many other types of data, proteins are composed of atoms and chemical bonds that possess inherent distance and angle information, rendering them unsuitable for placement in Euclidean domains. Geometrically, proteins are graphs by nature. This has sparked interests in the Graph Neural Networks (GNNs), which have emerged in recent years and excel in processing data from non-Euclidean domains. GNNs possess the ability to learn from the topology and connectivity of the graph structure, capturing complex relationships between atoms and their neighbors such as the distances between atoms, the angles between bonds, and the orientation of the entire molecule. These properties play a key role in determining the functions of the proteins and the effect of mutations on protein-protein interactions. Consequently, they can perform better for the processing of protein features and can provide more information than those are based on structural data only [8].

The advent of GNNs has come to the forefront in the past two years and has paved the way for the development of new learning and prediction methods that utilize these networks [8]. A study known as ScanNet [9] specifically introduces a spatio-chemical arrangement of neighbors neural network. GeoPPI [10], a geometric deep learning approach that generates the graph representations of mutations and learns the mapping from these representations to corresponding mutation effects, thereby enabling the prediction of affinity changes. Conspicuously, the test of GeoPPI [10] on the M1707 [7] dataset can reach a Pearson correlation of 0.74, which is the best performance of the benchmark methods on multi-point mutation datasets.

Furthermore, with the proliferation of large-scale pretrained language models in the broader field of machine learning, more and more large-scale pre-trained protein language models are emerging and being embraced in cheminformatics research. These include models that focus on capturing evolutionary signals within proteins, leveraging extensive sequence information to predict their structure, functions, and other foundational properties exclusively from their amino acid sequences, which is similar to using large language models to derive semantic information and language patterns from the contexts [11–14]. Typically developed through unsupervised strategies on sufficiently large datasets, these models exhibit intricate architectures with a large number of parameters, enabling them to encapsulate an extensive range of sequence information and complex feature interdependencies. As a result, they are well-equipped to generalize effectively across diverse branching tasks, marking a significant stride towards enhanced comprehension and application within the domain of protein-centric research. These findings serve as a compelling evidence of the potential of deep learning methods, particularly geometric ones, in accurately predicting the effects of the protein mutations on PPIs. Furthermore, they drive us to explore innovative avenues for model enhancement, integrating valuable insights from prior investigations to develop more effective approaches and achieve meaningful outcomes.

The present study successfully designs a novel method for predicting the effects of protein mutations on protein-protein interactions based on the three-dimensional structure of protein-protein complexes. The proposed model improves efficiency, accuracy and generalization ability over traditional models, thereby showing promise for applications in studying complex genetic diseases as well as in drug development.

Materials and methods

Datasets

In this work, we used five open-source datasets containing single-point and multi-point mutations of known proteins, with experimentally determined effects of these mutations on the folding free energy. Additionally, we incorporated a dataset for a case study, comprising potent antibodies targeting SARS-CoV-2 S protein complexes. For the five open-source datasets, the graph inputs for the neural network were derived from the experimentally resolved complex structures provided within the datasets. For the case study dataset, due to the absence of high-resolution structures, the graph inputs were generated from homology models built by the Rosetta3 tool based on sequence information. Rosetta3 was selected due to its foundation in physicochemical modeling and statistical mechanics, offering greater interpretability [15, 16].

The training and testing process involves the following five datasets (Table 1): S2648 [17], S3421, S4169 [18], M1101, and M1707 [7], which have been widely used to train and evaluate PPI prediction methods. These datasets include data on the changes in thermodynamic energy and kinetic rate constants upon mutations in PPIs, with the solved complex structures. S2648 encompasses a total of 2648 single-point mutations occurring in 131 distinct globular proteins, while S3421 comprises 3421 mutations identified experimentally in 150 proteins. Additionally, S4169 consists of 4169 variants selected from 319 different complexes, representing single-point mutations filtered from the SKEMPI 2.0 dataset [18]. The dataset denoted as M1101, on the other hand, includes 1101 distinct data points, which consist of both singlepoint mutations and multi-point mutations. The dataset M1707 comprises a total of 1337 variants, which include multi-point mutations along with their corresponding reversed mutations in certain regions. These data will be used as predictors of changes in binding affinity to assess the effect of PPIs.

Method

Problem formulation

A protein–protein complex arises from the non-covalent interaction between two or more protein molecules, facilitated through hydrogen bonds, electrostatic forces, and van der Waals forces [19]. This interaction can lead to structural changes in the protein conformation, which can ultimately impact its activity or function. The strength of the protein–protein interaction is often measured by binding affinity, which is estimated using the change in Gibbs free energy ($\Delta \Delta G$) associated with a mutation.

In this study, the main objective is to prognosticate the $\Delta\Delta G$ that is expected to occur in the protein–protein complex, because of mutations in the protein structure.

Table 1 Summary	of datasets
-----------------	-------------

Dataset	Data Type	Task	Number of proteins complexes	
S2648	Single Mutation	Regression	2648	
S3421	Single Mutation	Regression	3421	
S4169	Single Mutation	Regression	4169	
M1101	Single and Multipoint Mutation	Regression	1101	
M1707	Multipoint mutation	Regression	1707	

$$\Delta \Delta G = \Delta G_{wild-type} - \Delta G_{mutant}, \tag{1}$$

where ΔG represents the unfolding energy of a protein. $\Delta G_{wild-type}$ signifies the change in free energy upon binding in the wild-type protein complex, while ΔG_{mutant} represent the same quantity for the mutant complex.

Graph Initialization

The input data derived from the Protein Data Bank (PDB) files typically encompasses the elemental composition and structural details pertaining to protein complexes. To focus on the region where the mutation occurs and capture the salient feature information, some methods focus directly on mutant partial features, such as ProS GNN [20], which extract data pertaining to the mutant segment and utilizing message passing to effectively capture molecular characteristics. Drawing on this strategy, in this work we first "pruned" the involved and adjacent residues from the original protein-protein complex and its corresponding mutants, and then we devised the feature vector for individual atoms, encapsulating both vertex and edge information. This kind of feature vector effectively encoded a spectrum of attributes, including elemental identity, the number of neighboring atoms, implicit valence, and the presence of aromatic bonds. The "pruning" process not only facilitates targeted model training but also enhances computational efficiency. Furthermore, with these encoded features, we could successfully initialize the input graph for subsequent stages of analysis, wherein the nodes denote atoms, and the edges denote chemical bonds, thereby enabling a more focused and efficient exploration of molecular interactions.

Protein language model

Concurrently with the training and demonstration of the remarkable general-purpose language capabilities of large-scale language models, their transfer applications within the field of biochemistry have been evolving [21]. This evolution has given rise to the emergence of large-scale protein language models. Similar to their counterparts in the domain of natural language, protein language models possess the capability to deduce and acquire knowledge from large-scale data sets, have been regarded as their corpus, intrinsic biochemical attributes, diverse structural layers, and the implicit functional principles embedded within sequence information, which can be seen as their "contextual" relationships. These pretrained protein language models find utility in various downstream applications, including predicting protein structures, inferring protein functions, and generating novel sequences.

Furthermore, it is worthwhile to contemplate language models that place a particular emphasis on capturing

evolutionary information embedded within sequences [22]. It has been widely acknowledged that protein sequences in organisms are not randomly arranged permutations of amino acids, but rather exhibit discernible patterns, attributable to natural selection. For example, the pattern of variation of a protein in its family can reflect its structure, or non-independently evolved proteins may interact in a tertiary folded structure. Some methods that incorporate evolutionary profiles, such as SSIPe, have obtained good results [23]. Such patterns are presumably not captured by limited-scale models trained on small datasets.

Model structure

We propose a new architecture, named GES-PPI (a Graph-based neural network integrated with Evolutionary Scale modeling for Protein–Protein Interactions prediction), to predict the effects of mutations on protein–protein interactions The model consists of two primary components: a gated GNN and a graph transformer, as illustrated in Fig. 1. Overview of the GES_PPI Model Architecture. The pruned wild-type information and mutation information are fed into the model separately for analysis.

A gated GNN [24] is subsequently utilized to process the atomic features of the pruned region and facilitate the mapping of the protein's 3D structural information and component composition to a high-dimensional representation. Here we employ a graph convolutional network in which the key idea is to apply convolution over the graph with the propagation rule of

$$f(H^{(l)}, A) = \sigma \left(D^{-\frac{1}{2}} \hat{A} D^{-\frac{1}{2}} H^{(l)} W^{(l)} \right),$$
(2)

with the $\hat{A} = A + I$. $A \in \mathbb{R}^{n \times n}$ is the adjacent matrix, where *n* is the number of the mutant atoms. *I* is the identity matrix, and *D* is the diagonal node degree matrix of \hat{A} . With this rule, in each layer, the atom features $H^{l}(H^{l} \in \mathbb{R}^{n \times d})$, undergo successive iterations of graph convolution, leading to the generation of an updated feature set:

$$H^{(l+1)} = LeakyReLU(WAH^{l}).$$
(3)

with the LeakyReLU defined as:

Leaky ReLU(x) =
$$\begin{cases} 0.01 x, & \text{for } x < 0\\ x, & \text{for } x \ge 0 \end{cases}$$
(4)

 $W \in \mathbb{R}^{n \times n}$ is the weight matrix, and *d* denotes the dimension of the hidden state. To enhance the performance of feature extraction, we seamlessly incorporate the gating mechanism into the network. The gated graph



Fig. 1 Overview of the GES_PPI Model Architecture. The GES_PPI model combines structural and evolutionary insights for predicting $\Delta\Delta G$. Initially, the protein complexes are trimmed and processed by a gated GNN to capture local structural information. Subsequently, a Graph Transformer refines this representation by learning long-range relationships. With additional features derived from the pre-trained ESM model, the comprehensive feature set is used to predict the $\Delta\Delta G$, indicating the difference in stability between the wild-type and mutant protein complexes

layer is characterized by a linear combination of $H^{(l)}$ and $H^{(l+1)}$:

$$H_{gate} = GH^{(l)} + (1 - G)H^{(l+1)},$$
(5)

with

$$G = \sigma \left(W_{gate} \left[H^{(l)}, H^{(l+1)} \right] + B \right).$$
(6)

 $W_{gate} \in \mathbb{R}^{n \times d}$ should be the learnable weight matrix and $B \in \mathbb{R}^{n \times d}$ be the bias matrix. $\sigma(\cdot)$ is the sigmoid activation \cdot The gate connection would be added to the first layer and produce the final output $H_{gated_gcn}^{out} \in \mathbb{R}^{n \times d}$:

$$H_{gated_gcn}^{out} = H^1 + H_{gate}.$$
 (7)

During this step, each atom within the graph assimilates local information from its adjacent atoms and bonds, leading to the updating of its features. By aggregating the information from all atoms, global features could be obtained, and the total molecular energy could be calculated from the energy contributions of all individual atomic vectors in the generated feature vectors. Consequently, we obtained feature vectors for both the wild-type and mutant protein–protein complexes. To quantify the impact of mutations, we subtracted the characteristic vector of the mutant complex from that of the wild type, yielding a contrasting feature vector. The final change in binding free energy ($\Delta\Delta G$) is subsequently computed utilizing this contrasting vector, elucidating the energetic distinctions between the two complex states.

The present gated GNN sub-module has already exhibited an ability to acquire salient features and attributes inherent in the structure itself. Despite its proficiency, the gated GNN, similar to other GNNs, is heavily reliant on the links and adjacent features present in the graph. However, the determinants implicated in the protein complex mutations can be diverse and intricate, and even when the focus is solely on the mutated regions, it is imperative for the model to capture distant dependencies. This leads us to the transformer [25], which is built on the attention mechanism that transcends sequential relations and is unconstrained by links, consequently facilitating global inference capabilities.

In this work, we incorporate the widely adopted framework proposed by GraphTrans [26] in order to enhance the layer stacks of a single gated GNN and assist in representing long-range contextual relationships. The transformer sub-network employs herein functions as a distinctive readout module for the preceding gated GNN, facilitating pairwise interactions between learning graph nodes and subsequently amalgamating them into unique token embeddings. Specifically, after obtaining the final per-node GNN encodings, these representations are forwarded to the Transformer sub-network, which initially performs a linear projection of the per-node encodings into the Transformer dimension and subsequently conducts layer normalization to normalize the embeddings

$$H_{tf}^{0} = LN \left(W^{proj} H_{gatedgen}^{out} \right), \tag{8}$$

where $LN(\cdot)$ is the layer normalization function and $W^{proj} \in \mathbb{R}^{d_{tf} \times d_{tf}}$ is a learnable matrix. A transformer operation is performed on the projected embeddings and the embedding of each node is updated with

$$h_{\nu}^{l+1} = \sum_{w \in \mathcal{V}} \alpha_{\nu w}^{l+1} W_l^V h_w^l.$$
⁽⁹⁾

 α_{vw}^{l+1} is the attention value between node v and w, and $W_l^V \in \mathbb{R}^{d_{tf} \times d_{tf}}$ is the value matrix. The calculation of the attention matrix $Attn^{l+1} \in \mathbb{R}^{N \times N}$ between all pairwise nodes is performed as follows:

$$Attn^{l+1} = softmax\left(\frac{\left(W_{l+1}^{Q}H^{l}\right)^{T}\left(W_{l+1}^{K}H^{l}\right)}{\sqrt{d_{tf}}}\right), \qquad (10)$$

where W_{l+1}^Q , $W_{l+1}^K \in \mathbb{R}^{d_{tf} \times d_{tf}}$ denote the query and key matrices and *softmax*(·) is the softmax function.

The representation of the protein–protein complex derived from the graph transformer is obtained through a pooling operation. This resulting representation is then concatenated with the output of the evolutionary scale protein language model (ESM) model. This concatenation leads to the generation of the final representations denoted as F_{wild_type} and F_{mutant} , corresponding to the wild type complex and the mutant complex, respectively.

In order to further enhance the predictive efficacy of the proposed model, the integration of additional features derived from a pre-trained model are taken into consideration. In our approach, we harness the ESM, as shown in Fig. 1. The ESM is an unsupervised language model which is specifically designed to comprehend the evolutionary signals imprinted in protein sequences over vast timescales, yielding insights into the relationships between sequence, structure, and function [27, 28]. Trained by extensive sequence information available in protein databases, it is capable of extracting high-level features from protein sequences.

A protein language model operates by capturing intricate patterns and dependencies within protein sequences, enabling it to discern meaningful representations from amino acid sequences [14]. By incorporating ESM's global features into our model architecture, we enhance its ability to encapsulate nuanced information encompassing a protein's evolutionary context. These augmented features, synergistically combined with the localized features extracted through graph convolutions, enable a comprehensive representation of the protein– protein complexes. This integration empowers our model to capture both intricate local interactions and broader evolutionary trends, resulting in more accurate predictions of binding affinity changes. Through this innovative fusion of techniques, our approach underscores the significance of holistic feature extraction in advancing predictive modeling within the domain of protein–protein binding affinity prediction. Notably, by fusing the global features extracted from the ESM with the features obtained in this study after pooling them separately, more favorable results are obtained in the testing phase.

Model training

This study involves the development of a deep learning model designed to predict the changes in binding affinity stemming from mutations on protein-protein complexes. The model is developed using a supervised learning approach, where the input data consists of a set of protein-protein complexes, each with a known wildtype and mutant state, and the output consists of the predicted binding affinity changes due to the mutations.

To train the model, the input data is split into two sets: a training set and a test set, in an 8:2 ratio. We also performed model selection through a tenfold cross-validation (CV) on the training set. The training set is used to iteratively update the model parameters by minimizing the mean squared error between the predicted and actual binding affinity changes.

The tenfold CV is used to evaluate model performance during training and mitigate the risk of overfitting. The model is stopped after a fixed number of epochs, which is determined by monitoring the performance of the model during the cross-validation. The model architecture consists of a gated Graph Neural Network submodule, with a vertex vector dimension of 120, fully connected layers with a dimension of 1024, and Leaky ReLU as the activation function. The training process involves a batch size of 32, a dropout rate of 0.5, a learning rate of 0.001, and the Adam optimizer. The final model is then evaluated on an independent test set to assess its ability to generalize unseen data.

Model evaluation and baseline models

The present study is executed utilizing an Nvidia Geforce GTX 3070 Ti GPU, and the model is developed through the implementation in Pytorch. The primary criteria employed for assessing the accuracy of the predictions are the Pearson correlation coefficient (R_p) and the Root Mean Square Error (RMSE) of the experimental and predicted $\Delta\Delta G$ values.

The R_p , ranging from -1 to 1, qualifies both the intensity and direction of the relationship between two variables. When one variable changes, the other variable also changes in concordance with that direction. The RMSE is the difference between values predicted by the model and actual values, and it is calculated as the square root of the mean of the squared differences between the predicted $\Delta\Delta G$ values and actual $\Delta\Delta G$ values. The formula for RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_{pred}(i) - y_{true}(i))^2},$$
 (11)

where n is the total number of samples, $y_{pred}(i)$ is the predicted $\Delta\Delta G$ value for the *i*-th sample, and $y_{true}(i)$ is the actual $\Delta\Delta G$ value for the *i*-th sample.

We perform comparative evaluations of GES_PPI against 4 baseline methods. These include two state-of-the-art geometric approaches: (1) a supervised gated_GNN with an input trimming strategy (ProS_GNN (20)); (2) a gradient boosting tree wherein the input graph is generated by a self-supervised perturbation-based geometric encoder (GeoPPI (10)). Besides, we compare our model with other two well-known studies in similar topic: a random forest method (MutaBind2 (7)) and a topology-based GBT architecture (TopGBT (5)), to demonstrate the power of our proposed approach.

The training dataset is utilized to iteratively update the model parameters by minimizing the RMSE between predicted and actual binding affinity changes. We test the model and its baselines in its entirety and without the inclusion of the ESM pre-trained model on each of the five test datasets. Finally, there are independent test sets to evaluate the model's ability to generalize to unseen data.

Ablation analysis

In addition to the effects of the previously mentioned ESM components on the expressiveness of the model, to investigate the contribution of the pre-trained model, pruning strategy, and transformer framework respectively in our proposed method, we conduct ablation analysis experiments on the S2648 [17] dataset. We sequentially remove each significant component from the model and evaluate its performance on the prediction of $\Delta\Delta G$. First, we remove the graph transformer component and only use the gated GNN to process the atomic features of the pruned region. We then evaluate the impact

of the "pruning" step by using the whole graph of the protein complex.

Case study

In this section, we assess the realistic utility of our framework employing SARS-CoV-2 as an illustrative case study [29]. We seek to investigate our model's applicability in capturing the effect of antibody (Abs) mutations on SARS-CoV-2 binding affinity [30, 31]. For this evaluation, we use a test dataset containing potent Abs to SARS-CoV-2 S protein complexes [10, 29], primarily identified from recovering patients affected by SARS-CoV-2. For each Ab, we identified templates with high sequence homology and used the "comparative modeling" feature in Rosetta3 to generate possible structures. The structure with the highest score was then selected.

The performance of GES_PPI is rigorously assessed by measuring the disparity between the predicted and experimental binding affinity changes of each pair of structurally similar antibodies to SARS-CoV-2. This investigation provides valuable insights into the effectiveness of GES_PPI in accurately predicting the impact of antibody mutations on the binding affinity to SARS-CoV-2, further extending its potential applications in addressing critical challenges posed by the ongoing pandemic.

Results

Model performance

The proposed model was first evaluated on three single-mutation datasets. As shown in Table 2, GES_PPI garnered the highest correlation coefficients and the lowest RMSEs across all these three datasets. Furthermore, it also exhibited good performance on the two datasets containing multi-point mutations, with highest correlation coefficients on M1101 and lowest RMSE on M1707. It is worth noting that GES_PPI outperformed the baseline model, ProS_GNN, by improving the correlation by 6% on M1101, which contains both single-point and multi-point mutations.

To demonstrate the ability of the proposed method itself to extract and analyze features, we removed the ESM module for a complete test. As illustrated in Table 2, the model (called gnn_PPI) could still exhibit higher correlation coefficients than benchmarks on both the S2648 [17] and S3421 datasets, as well as lower RMSE on the S3421 and S4169 [18] datasets. When tested with the S3421 dataset, the proposed method

Method	Single Mutation						Multi-Point Mutation			
	S2648		\$3421		S4169		M1101		M1707	
	R _p	RMSE	R _p	RMSE	R _p	RMSE	R _p	RMSE	R _p	RMSE
GES_PPI	0.649	1.132	0.717	1.641	0.689	1.563	0.569	1.762	0.754	2.142
gnn_PPI*	0.627	1.220	0.673	1.694	0.625	1.570	0.543	1.761	0.740	2.176
ProS_GNN	0.608	1.207	0.663	1.724	0.657	1.619	0.510	2.024	0.694	2.343
GeoPPI	0.584	1.188	0.682	1.701	0.592	1.588	0.562	1.885	0.727	2.235
MutaBind2	0.532	1.228	0.690	1.695	0.628	1.596	0.529	1.794	0.712	2.256
TopGBT	0.451	1.352	0.543	1.846	0.409	1.628	-	-	-	-

Table 2 Comparison of the proposed method with pre-trained protein language model for the single and multi-point mutations in terms of R_P and RMSE

The bold values represent the best results within each column, corresponding to the specific dataset and evaluation metric

-: the dash sign indicates the results of the corresponding methods are not available

*: gnn_PPI represents the method without the ESM pre-trained model

achieved Rp = 0.717 and RMSE = 1.641 (Fig. 2a). These findings suggest that the proposed method performs competitively on datasets of different sizes. In addition, the test results on multi-point mutation datasets demonstrated that the proposed model achieved the highest correlation coefficients and the lowest RMSEs on both datasets. When tested with the M1707 [7] dataset, the proposed model achieved Rp = 0.754 and RMSE = 2.142(Fig. 2b).

To further evaluate the robustness of the high correlation of our proposed models, we calculated the standard deviation of the correlation values based on tenfold cross-validation in the training set, which is shown in Table 3. Overall, we can see the correlation values are relatively stable, and their standard deviations are relatively small.

What is more, computational efficiency is also an important consideration in the development of this predictive model, especially in the context of highthroughput screening and drug discovery. In this study, we compared the inference time of the five benchmark models, as illustrated in Fig. 3, in order to assess the computational speed and efficiency of our proposed approach. The results indicate that the proposed GES_ PPI model exhibits significantly faster prediction times compared to most of the benchmarks, with an average time of 16 s for predicting the binding affinity change of a single mutant. Although the computational time for the proposed model's prediction is one second slower than that of ProS_GNN [20], the experimental results presented above show that GES PPI exhibits better predictive performance. This is noteworthy considering that the original purpose of selecting



Fig. 2 Predicting $\Delta\Delta G$ for mutations in S3421 and M1701 datasets respectively

Table 3 The results of the R_P and RMSE with standard deviation of the proposed models (GES_PPI and gnn_PPI)

Method	Dataset	R _p	RMSE
GES_PPI	S2648	0.6491 ± 0.0108	1.1324 ± 0.0114
	S3421	0.7166 ± 0.1022	1.6413 ± 0.0840
	S4169	0.6892 ± 0.0126	1.5633 ± 0.0143
	M1101	0.5679 ± 0.0221	1.7622 ± 0.1421
	M1707	0.7538 ± 0.1097	2.1421 ± 0.0988
gnn_PPI*	S2648	0.6272 ± 0.0441	1.2197 ± 0.0308
	S3421	0.6725 ± 0.1320	1.6944 ± 0.0806
	S4169	0.6254 ± 0.1052	1.5697 ± 0.1231
	M1101	0.5433 ± 0.2597	1.7607 ± 0.2080
	M1707	0.7397 ± 0.1956	2.1756 ± 0.1427

*: gnn_PPI represents the method without the ESM pre-trained model



for the prediction of individual methods

automated methods is to enable time-saving and efficient prediction.

Ablation analysis

The results of the ablation analysis experiments are presented in Table 4, indicating that the performance of the model was noticeably affected by removing any part of the model. The performance is represented by mean \pm standard deviation of the evaluation metric across tenfold CV. The results emphasize the importance of targeted training and the graph transformer. Without either of these two components, the prediction error of the model rises. The performance of the model also dropped from GES_PPI to gnn_PPI, indicating that the features from large-scale pre-trained model are essential components for accurate $\Delta\Delta G$ prediction. In general, the ablation analysis results demonstrate that each component of our proposed method plays a crucial role in predicting $\Delta\Delta G$.

This gives the test results of model with different components.

Case study

The Abs dataset, with an average number of mutations around 11, is more complex than the multi-point mutation dataset, M1707 [6] (with an average number of mutations around 3), used for training and consequently poses a significant prediction difficulty. It is noteworthy that in such a case, GES_PPI still achieved a correlation result of 0.63, a significant advantage over Mutabind2 (0.29 correlation) (Fig. 4), which was also tested using this Abs dataset. This suggests that GES_PPI, while primarily designed for general PPIs, has the potential to extend its applicability to more specialized and challenging contexts.

Interpretability

Beyond the evaluation of a model's performance, it holds significance to delve into the identification of specific subcomponents within the complex structure that exert a more pronounced influence on predictions. The attention mechanism employed in Transformers facilitates the identification of potential biomarkers or critical regions within the protein structure, which represents an insight into the contribution of different substructures of the protein structure by visualizing its attentional weight in this part of the network., and thus

Table 4 Results of ablation analysis experiments

Components				Performance		
Pruning	Gated_GCN	Transformer	ESM	RMSE	R _p	
$\overline{\checkmark}$	\checkmark	\checkmark	\checkmark	1.1324±0.0114	0.6491 ± 0.0108	
\checkmark	\checkmark	\checkmark		1.2197 ± 0.0308	0.6272 ± 0.0441	
\checkmark	\checkmark			1.2305 ± 0.0327	0.5818 ± 0.0319	
	\checkmark			1.2516 ± 0.0454	0.4997 ± 0.0336	

The bold values highlight the RMSE or Rp values themselves, distinguishing them from the standard deviation values after the ± symbol



Fig. 4 Case study on potent Abs to SARS-CoV-2 S protein complexes. The plot shows pairwise prediction performance between structurally similar Abs

plays a crucial role in enhancing the interpretability of our model's predictions.

We took the reduced form of DsbA from Escherichia coli, the 1A23 protein, as an example, for which the wild-type structure is shown in Fig. 5a. We selected the amino acid at position 31, where the mutation occurs from Histidine (H) to Tyrosine (Y). We extracted the attention weights in the Graph Transformer sub-network and drew the attention matrix, as shown in Fig. 5b. Following this, we correlated the extrated attention weights with the initial feature matrix for wild-type data and colored the selected wild-type structure in the figure. As shown in Fig. 5a, the brighter areas indicate that, in the upcoming

mutation, will contribute more in the calculation of energy change. Although the mutation occurs from one type of amino acid to another type, the change in sidechain composition and structure affects how the amino acid interacts with other molecules, and may therefore be captured as different contributions, which are the weights, in model learning.

By enabling the model to focus on relevant features and interactions within the input data, we can gain valuable insights into the influential regions of the protein structure for the prediction task. This interpretability analysis empowers researchers and domain experts to identify key molecular interactions and structural elements that significantly contribute to changes in binding affinity.

Discussion

The results of this study are subject to certain limitations and potential sources of bias that are likely to impact their validity. Specifically, one such limitation arises from the scarcity of thermodynamic measurements of proteins in the current databases, which may restrict the model's applicability to larger and more diverse datasets. The accuracy of binding affinity predictions may be sensitive to the quality of the 3D structures used, and variations in $\Delta\Delta G$ values among individual datasets due to diverse experimental conditions may further impact the assessment and veracity of the predicted results [33–35].

The limitations in the diversity of mutations across existing real datasets might also pose challenges in extending the model to datasets with greater variation,



Fig. 5 Visualization of attention weights for the wild-type structure of the 31st amino acid in the 1A23 protein chain. **a** Visualization of the wild-type structure of the 1A23 protein at position 31, highlighting the atom significance using attention weights. **b** Attention matrix showing the significance of atoms when a mutation occurs in the wild-type structure

potentially impacting its practical utility [33]. Alanine scanning, for example, involves the systematic substitution of amino acids with alanine to determine their role in protein function and interactions. While this approach is useful for identifying key residues, it may introduce biases as it primarily features alanine substitutions. This limited diversity may not reflect the full range of amino acid substitutions that may occur in the natural environment. Furthermore, potential similarities within the datasets may impact prediction results.

We evaluated the similarity between training and test datasets by aligning protein sequences using Protein BLAST, focusing on the mutated residue and its adjacent residues (See Supplementary Materials for details). Similarity distributions for both pruned and full sequences are shown in Figures S1 and S2. To assess the impact of sequence similarity on prediction, we applied thresholds (1%, 5%, 10%) and iteratively removed test samples with high similarity to the training set. The effect of this filtering on dataset size and prediction performance is summarized in Tables S1 and S2. Although removing highly similar sequences slightly reduced accuracy, the model's performance, as indicated by the Pearson correlation coefficient (Rp), remained comparable to or better than baseline methods (Table 2). However, analyzing protein similarity is inherently complex, as some protein complexes may exhibit significant differences in sequence or structure while maintaining functional similarities, often due to remote homology. Addressing these complexities will require advanced methodologies capable of capturing the intricate dynamics of protein-protein interactions (PPIs), as well as the development of more comprehensive and diverse datasets.

Protein-protein interactions are inherently complex processes that encompass a diverse array of causes, consequences, and impacts. Our proposed model aims to predict the changes in free energy associated with protein mutations and their significant effects on PPIs. However, it may not comprehensively capture the full range of PPI dynamics. The limitations in publicly available wet lab data make it challenging to create large-scale, systematic pre-training models that could fully address these complexities. Future advancements will require extensive datasets and improved methodologies to better understand the intricate relationships within PPIs.

Despite these limitations, the results of this study highlight the model's strong generalizability, as evidenced by its robust performance on both single-point mutation datasets and multi-point mutation datasets. These findings have significant implications for protein engineering and drug discovery, and the proposed model holds promise for diverse settings and applications in the field of protein–protein binding affinity prediction. While careful consideration of the limitations is warranted, the study's outcomes can provide valuable insights and contribute to advancing research in the field of proteinprotein interactions. Moreover, through the analysis of attention weights, we can highlight specific residues, binding sites, or interaction patterns that exert a strong influence on the model's predictions [25]. This interpretability aspect not only enhances our understanding of the underlying mechanisms governing protein-protein interactions but also provides essential insights for guiding further investigations and experimental validations. The attention mechanism highlights the specific regions and interactions driving the model's predictions, making it a valuable tool for unraveling the complexities of protein-protein interactions. This will continue contributing to the interpretability and transparency of this kind of model, both in this work and future studies.

In terms of future work, we may expand our focus to the whole protein structure to pinpoint every amino acid or specific atom within the protein that are influential in determining the binding affinity change, not only improving the calculation accuracy. Moreover, the test results show room for further improvement in predicting the effects of multi-point mutations, suggesting that further development may lead to better accuracy in these complex scenarios. Additionally, the proposed model could be improved by incorporating larger and more diverse datasets or utilizing a variety of data sources, which can potentially enhance its performance. Exploring different types of features and models is also a potential way for optimizing the performance. It is also recommended to explore other graph representation learning techniques to discover more efficient methods for automation.

Conclusions

The present study develops a large language modeldriven graph neural network model to predict the effect of mutation on protein-protein interaction binding affinity. To streamline the training process and emphasize the characteristics of the mutant component, the mutant segment of the protein complex was initially extracted. This model incorporates a gated graph neural network to capture atomic-level features and a graph transformer for embedding projection, ultimately resulting in the prediction of $\Delta\Delta G$. The efficacy of the model was subjected to comprehensive experimental evaluation, with results demonstrating its robust competitiveness on baseline models using five datasets. This novel approach to the study of protein stability alterations through the implementation of GNNs carries significant implications for future stability prediction endeavors.

Abbreviations

RMSE Root mean square error

- GNN Graph neural network
- PPI Protein-protein interaction
- CNN Convolutional neural network
- ESM Evolutionary scale protein language model
- Abs Antibody
- PDB Protein Data Bank
- CV Cross-Validation

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13321-025-00979-5.

Additional file 1.

Author contributions

Conceptualization: PH, CZ, YS. Data curation: CZ, YS. Methodology: PH, CZ, YS. Data analysis: CZ. Validation: CZ. Software: CZ. Supervision: PH, YS. Funding acquisition: PH. Initial draft: CZ. Final manuscript: PH, CZ, YS. All authors read and approved the final manuscript.

Funding

This work was supported in part by the Canada Research Chairs Tier II Program (CRC-2021–00482), the Canadian Institutes of Health Research (PLL 185683, PJT 190272), the Natural Sciences and Engineering Research Council of Canada (RGPIN-2021–04072) and The Canada Foundation for Innovation (CFI) John R. Evans Leaders Fund (JELF) program (#43481).

Data availability

Data is provided within the manuscript. Other data sets and the code can be found in our GitHub repository: https://github.com/Caiya-Zhang/GES_PPI_for_binding_affinity_prediction.

Code availability

The code and datasets can be found in our GitHub repository: https://github. com/Caiya-Zhang/GES_PPI_for_binding_affinity_prediction.

Declarations

Completing interests

The authors declare no competing interests.

Author details

¹Department of Computer Science, Western University, London, ON, Canada.
²Department of Computer Science, University of Manitoba, Winnipeg, MB, Canada.
³Department of Biochemistry, Western University, London, ON, Canada.
⁴Department of Oncology, Western University, London, ON, Canada.
⁵Department of Epidemiology and Biostatistics, Western University, London, ON, CN, ON, Canada.
⁶The Children's Health Research Institute, Lawson Health Research Institute, London, ON, Canada.

Received: 31 March 2024 Accepted: 27 February 2025 Published online: 21 March 2025

References

- Stefl S, Nishi H, Petukh M, Panchenko AR, Alexov E (2013) Molecular mechanisms of disease-causing missense mutations. J Mol Biol 425:3919–3936
- Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 33:W306
- Sun T, Zhou B, Lai L, Pei J (2017) Sequence-based prediction of protein protein interaction using a deep-learning algorithm. BMC Bioinf 18(1):277
- Yang F, Fan K, Song D, Lin H (2020) Graph-based prediction of proteinprotein interactions with attributed signed graph embedding. BMC Bioinf 21(1):323

- Wang M, Cang Z, Wei G-W (2020) A topology-based network tree for the prediction of protein-protein binding affinity changes following mutation. Nat Mach Intell 2:116–123
- Rodrigues CH, Myung Y, Pires DE, Ascher DB (2019) MCSM-PPI2: predicting the effects of mutations on protein-protein interactions. Nucleic Acids Res 47:W338
- Zhang N, Chen Y, Lu H, Zhao F, Alvarez RV, Goncearenco A, Panchenko AR, Li M (2020) Mutabind2: predicting the impacts of single and multiple mutations on protein-protein interactions. iScience 23:100939
- Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS (2021) A Comprehensive Survey on Graph Neural Networks. IEEE Trans Neural Netw Learn Syst 32:4–24
- Tubiana J, Schneidman-Duhovny D, Wolfson HJ (2022) Scannet: an interpretable geometric deep learning model for structure-based protein binding site prediction. Nat Methods 19:730–739
- Liu X, Luo Y, Li P, Song S, Peng J (2021) Deep geometric representations for modeling effects of mutations on protein-protein binding affinity. PLOS Comput Biol 17:e1009284
- 11. Wang S-W, Bitbol A-F, Wingreen NS (2019) Revealing evolutionary constraints on proteins through sequence analysis. PLOS Comput Biol 15:e1007010
- Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, Qin C, Žídek A, Nelson AW, Bridgland A et al (2020) Improved protein structure prediction using potentials from deep learning. Nature 577:706–710
- Min B, Ross H, Sulem E, Veyseh AP, Nguyen TH, Sainz O, Agirre E, Heintz I, Roth D (2023) Recent advances in natural language processing via large pre-trained language models: a survey. ACM Comput Surv 56:1–40
- Madani A, Krause B, Greene ER, Subramanian S, Mohr BP, Holton JM, Olmos JL, Xiong C, Sun ZZ, Socher R et al (2023) Large language models generate functional protein sequences across diverse families. Nat Biotechnol 41:1099–1106
- Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman KW, Renfrew PD, Smith CA, et al. ROSETTA3. Computer Methods, Part C 2011, 545–574.
- Leman JK, Weitzner BD, Lewis SM, Adolf-Bryfogle J, Alam N, Alford RF, Aprahamian M, Baker D, Barlow KA, Barth P et al (2020) Macromolecular modeling and design in rosetta: recent methods and frameworks. Nat Methods 17:665–680
- Dehouck Y, Grosfils A, Folch B, Gilis D, Bogaerts P, Rooman M (2009) Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: Popmusic-2.0. Bioinformatics 25:2537–2543
- Jankauskaitė J, Jiménez-García B, Dapkūnas J, Fernández-Recio J, Moal IH (2019) SKEMPI 2.0: An updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. Bioinformatics 35(3):462–469
- 19. Kuo LC (2011) Fragment-based drug design: tools, practical approaches, and examples. Elsevier/Academic Press, San Diego
- 20. Wang S, Tang H, Shan P, Zuo L (2023) Pros-GNN: Predicting effects of mutations on protein stability using graph neural networks. Comput Biol Chem 107:107952
- Fang Y, Liang X, Zhang N, Liu K, Huang R, Chen Z, Fan X, Chen H (2024) Mol-instructions: a large-scale biomolecular instruction dataset for large language models. https://doi.org/10.48550/arXiv.2306.08018
- Hie BL, Shanker VR, Xu D, Bruun TU, Weidenbacher PA, Tang S, Wu W, Pak JE, Kim PS (2023) Efficient evolution of human antibodies from general protein language models. Nat Biotechnol 42:275–283
- Huang X, Zheng W, Pearce R, Zhang Y (2019) SSIPE: accurately estimating protein-protein binding affinity change upon mutations using evolutionary profiles in combination with an optimized physical energy function. Bioinformatics 36:2429–2437
- 24. Li Y, Tarlow D, Brockschmidt M, Zemel R (2015) Gated graph sequence neural networks. arXiv:1511.05493
- 25. Lim J, Ryu S, Park K, Choe YJ, Ham J, Kim WY (2019) Predicting drug-target interaction using a novel graph neural network with 3d structureembedded graph representation. J Chem Inf Model 59:3981–3988
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17) 2017.

- Wu Z, Jain P, Wright M, Mirhoseini A, Gonzalez JE, Stoica I. Representing long-range context for graph neural networks with global attention. NeurIPS 2021.
- Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, Guo D, Ott M, Zitnick CL et al (2021) Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci USA 118(15)
- 29. Flores SC, Alexiou A, Glaros A (2021) Mining the protein data bank to improve prediction of changes in protein-protein binding. Plos One 18(2):e0281259
- Barnes CO, Jette CA, Abernathy ME, Dam K-MA, Esswein SR, Gristick HB, Malyutin AG, Sharaf NG, Huey-Tubman KE, Lee YE et al (2020) SARS-COV-2 neutralizing antibody structures inform therapeutic strategies. Nature 588:682–687
- Zhang J, Xiao T, Cai Y, Lavine CL, Peng H, Zhu H, Anand K, Tong P, Gautam A, Mayer ML et al (2021) Membrane fusion and immune evasion by the spike protein of SARS-COV-2 delta variant. Science 374:1353–1360
- 32. Chen J, Gao K, Wang R, Wei G-W (2021) Revealing the threat of emerging SARS-COV-2 mutations to antibody therapies. J Mol Biol 433:167155
- Schirra HJ, Renner C, Czisch M, Huber-Wunderlich M, Holak TA, Glockshuber R (1998) Structure of reduced dsba from escherichia coli in solution. Biochemistry 37:6263–6276
- 34. Sanavia T, Birolo G, Montanucci L, Turina P, Capriotti E, Fariselli P (2020) Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine. Comput Struct Biotechnol J 18:1968–1979
- Liu Z, Pan W, Zhen X, Liang J, Cai W, Yuan K, Lin GN. Will Alphafold2 Be Helpful in Improving the Accuracy of Single-Sequence PPI Site Prediction? 2022 10th International Conference on Bioinformatics and Computational Biology (ICBCB) 2022.
- Geng C, Xue LC, Roel-Touris J, Bonvin AM (2019) Finding the ΔΔg spot: are predictors of binding affinity changes upon mutations in protein– protein interactions ready for it? WIREs Comput Mol Sci 9:e1410

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.