Pair-EGRET: Enhancing the prediction of protein-protein interaction sites through graph attention networks and protein language models

Ramisa Alam[1]
1705004@ugard.cse.buet.ac.bd

Sazan Mahbub[1,2]
smahbub@cs.cmu.edu

Md. Shamsuzzoha Bayzid [1,*]
shams_bayzid@cse.buet.ac.bd

[1]Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh
[2]Computational Biology Department, School of Computer Science
Carnegie Mellon University, Pittsburgh, PA 15213, USA
*To whom correspondence should be addressed

Abstract

Proteins are responsible for most biological functions, many of which require the interaction of more than one protein molecule. However, predicting protein-protein interaction (PPI) sites (the interfacial residues of a protein that interact with other protein molecules) remains a challenge. The growing demand and cost associated with the reliable identification of PPI sites using conventional experimental methods call for computational tools for automated prediction and understanding of PPIs. Here, we present Pair-EGRET, an edge-aggregated graph attention network that leverages the features extracted from pre-trained transformer-like models to accurately predict pairwise protein-protein interaction sites. Pair-EGRET works on a k-nearest neighbor graph, representing the three-dimensional structure of a protein, and utilizes the cross-attention mechanism on top of a siamese network to accurately identify interfacial residues of a pair of proteins. Through an extensive evaluation study using a diverse array of experimental data, evaluation metrics, and case studies on representative protein sequences, we find that our method outperforms other state-of-the-art methods for predicting PPI sites. Moreover, Pair-EGRET can provide interpretable insights from the learned cross-attention matrix. Pair-EGRET is freely available at `https://github.com/1705004/Pair-EGRET`.

1 Introduction

Proteins play a fundamental role in various cellular processes, often forming complexes through interactions between multiple proteins [1]. A thorough understanding of protein interfaces and the interacting residues involved is crucial in fields like disease research and drug development[2–7].

Protein-protein interaction sites (PPIS) are the residues of a protein that interact with residues from other proteins to form an interface between them. The problem of identifying interacting residues of proteins has two forms: i) *Partner-independent* prediction involves finding residues in an isolated protein that may interact with residues from any other protein [8, 9]. ii) The second form – and the one addressed in this study – is *partner-specific* prediction which involves identifying interacting residues of a protein for a specific partner protein. Partner-specific methods can further be classified into those that seek to only identify which residues constitute the interface between protein pairs (henceforth referred to as interface region prediction methods) and those that seek to identify specific

Machine Learning for Structural Biology Workshop, NeurIPS 2023.

pairs of residues (one from each protein) that interact with one another (henceforth referred to as pairwise PPIS prediction methods) [10, 11]. The latter is notably more challenging. This research addresses both forms, presenting an innovative method for accurately predicting interface regions and pairwise PPIS in protein complexes.

Experimentally identifying PPIS using wet lab methods is time-consuming and costly, resulting in the rise of various computational approaches as alternatives. Computational methods like protein-protein docking models [12–15], template-based methods [16–18] and machine-learning methods [19, 11, 20–24] employ different techniques like computationally rotating and translating proteins to produce different poses, comparing an unknown protein with a known query protein, and training models to learn features of interacting residues, etc. However, these methods suffer from numerous shortcomings including limited coverage, inability to predict novel interactions, and challenges in feature selection.

Some recent deep learning-based methods [10, 8, 25, 26] have demonstrated that incorporating information from both the primary amino-acid sequence and the 3D structure of a protein leads to more accurate identification of PPIS. For extracting features from the primary sequence, the recently developed protein language models [27, 28], trained on large datasets of 1D amino acid sequences, have been proved to be effective [8, 26, 29]. For encoding the 3D structural information of proteins, several geometric structures have been proposed [9, 30] among which Graph Neural Networks (GNN) [31] has proved to be useful for a number of methods [10, 25, 8]. GNNs have the capability to learn global and local contextual features for a residue from its neighborhood. Mahbub and Bayzid [8] proposed an edge-aggregated graph attention network [32] - a variant of GNN, in their model EGRET where both node-level and edge-level features are used for calculating the attention scores allowing the model to utilize the rich structural data encoded in the edges of the graph. EGRET was proposed for the task of partner-independent prediction of interaction sites.

In this study, building on the recent successful application of transformers and GAT networks in partner-independent PPI site prediction, we propose a novel deep learning model Pair-EGRET that extends the architecture of EGRET for both pairwise PPI site and interface region predictions. Pair-EGRET integrates GNN and protein language models to effectively leverage both structural and sequence-based information. Moreover, to allow each protein to attend to the relevant features from the other protein's residues, we adopted the concept of the cross-attention [33], widely used in Natural Language Processing (NLP) for incorporating information from multiple input sources or contexts. The combination of using transfer learning, edge-aggregated GAT networks, and cross-attention modules led to the improvement of Pair-EGRET compared to the best alternative methods on widely accepted benchmark datasets for partner-specific interaction prediction (both PPI site and interface region predictions). We have included case studies that visually inspect the predicted binding residues. We further visualize and interpret the representations learned by Pair-EGRET.

2 Approach

In our study, we represent the three-dimensional structures of the receptor and the ligand-protein of a complex, using directed k-nearest neighbor graphs [34]. Each graph node corresponds to an amino acid residue and is connected to its k closest neighbors via directed edges. Each node of the graph is characterized by some features of the corresponding residue. We used embedding vectors generated by ProtBERT [27] and some physicochemical properties of amino acids as node-level features. The distance and angle between residues served as the edge-level features for the weighted edges connecting them. For a detailed discussion of the graph representation of protein for Pair-EGRET, refer to Appendix A.

2.1 Architecture of Pair-EGRET

The architecture of Pair-EGRET can be discussed in two parts: (i) the architecture of EGRET [8], which was initially proposed for identifying interaction sites from a single protein, and (ii) our proposed extension to the EGRET architecture for predicting interaction sites between pairs of proteins. We will discuss the components of these two architectures briefly in this section. We strongly recommend reading through Appendix A for an in-depth understanding of both the original EGRET and the Pair-EGRET model.

2.1.1 Architecture of EGRET

In our study, we utilize the EGRET model as the foundation for our Pair-EGRET framework. Fig. A1 shows the end-to-end pipeline of EGRET with three core components:

i) The local feature extractor: This layer captures the local interactions between neighboring residues using a one-dimensional CNN applied to the node-level features.

ii) The edge-aggregated graph attention layer: This layer encodes the three-dimensional structural information of the neighborhood of each node. It employs an improved graph attention [32] mechanism, incorporating both node and edge-level features for calculating attention scores and applying aggregation – a process used in different GNN architectures [35, 32].

iii) The node-level classifier: This final layer linearly transforms the aggregated features obtained from the previous layer and applies sigmoid activation to generate interaction probabilities for each residue of a sequence.

2.1.2 Extension to EGRET for pairwise prediction

For solving both forms of partner-specific protein interaction prediction problem, we extend the EGRET model. We leverage the EGRET architecture to extract useful features from receptor and ligand proteins separately and add additional layers that combine features from both proteins and produce our desired outputs. There are five core components in the Pair-EGRET architecture:

i) Siamese EGRET network: Our proposed architecture of Pair-EGRET begins with a Siamese network, which consists of a pair of identical networks containing the first two modules of EGRET (local feature extractor and graph-based encoder). These identical networks share weights with one another, enabling the Siamese network to learn a common representation of both the ligand and the receptor.

ii) Positional encoder: The positional encoder module - inspired by Vaswani et al. [33], is used to utilize the information about the position of residues in a protein sequence. This module calculates a positional embedding for each residue and adds it to the feature representations obtained from the Siamese network. This can be useful because the location of a residue can impact its interaction with residues present in the partner protein in the complex.

iii) Multi-headed cross-attention layer: The multi-headed cross-attention layer employs the cross-attention mechanism [33] to combine features originating from both receptor and ligand proteins. Through an encoder-decoder structure, this layer facilitates the mutual exchange of information between the proteins. It quantifies the influence of a ligand residue on a receptor residue and vice versa by generating attention scores that help identify relevant residue pairs within a complex. We employ multiple attention heads within this module to capture various aspects of the input features.

iv) Pairwise classifier: The pairwise classifier layer in Pair-EGRET produces the final output for our first task: pairwise PPIS prediction. Within this module, a series of fully connected layers and activation functions are applied to a feature vector formed by combining the output of the previous layer for every possible ligand-receptor residue pair within the complex. The resulting output probability scores signify the likelihood of interaction between each of these residue pairs.

v) Interface region classifier: The interface region classifier generates the final output for our second task - identifying the interface region of protein complexes. Similar to the pairwise classifier, this layer also incorporates fully connected layers and activation functions to transform the feature representation of each individual protein into probability scores indicating the likelihood of each residue of the protein being a part of its interface. Ideally, any residue of a protein that interacts with any other residue of its partner protein should be identified as an interface residue by this layer.

Fig. 1 shows the overall end-to-end pipeline of Pair-EGRET where the first three modules are connected sequentially and are common to both forms of the partner-specific interaction prediction problem we are addressing. The final two layers are parallel networks that generate the outputs corresponding to each problem.

3 Experimental studies

3.1 Dataset

We evaluated Pair-EGRET on three benchmark datasets: i) Docking Benchmark version 5.0 [36] (DBD5), ii) Dockground X-ray unbound docking benchmark version 4 [37], and iii) A subset of

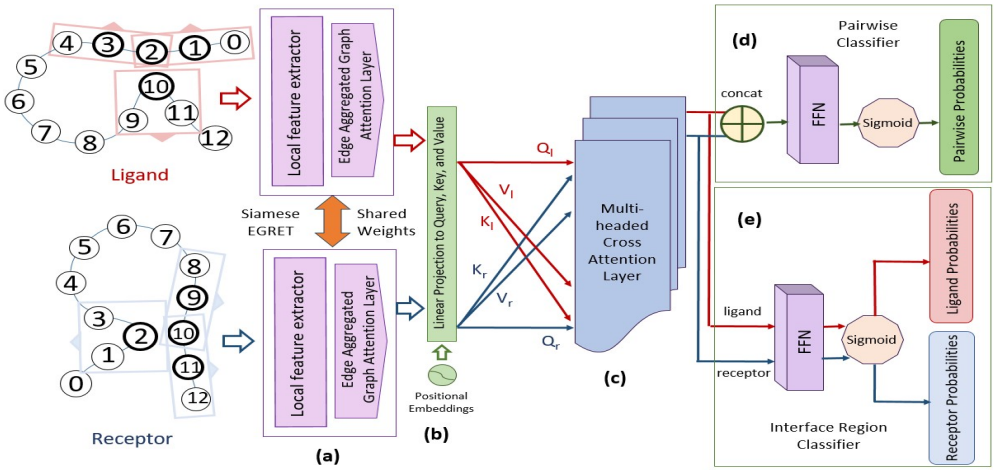


Figure 1: Schematic diagram of the overall pipeline of Pair-EGRET being applied to a receptor protein (r) and a ligand-protein (l). **(a)** Siamese EGRET network with shared weights. **(b)** Positional encoder being added to the individual proteins and linearly projected to three separate feature spaces– query, key, and value (discussed in Section A.3(iii)). **(c)** Both proteins transformed through the multi-headed cross-attention layer using the query vector of itself and the key and value vector of the other protein. **(d)** Pairwise classifier being applied to the merged features. **(e)** Interface region classifier being applied to individual proteins.

the MaSIF [9] dataset. Table B1 contains a summary of these datasets. Similar to prior work [10, 38, 11, 25, 25], we considered residue pairs to be interacting if they had non-hydrogen atoms within 6Å distance. We downsampled the non-interactive residue pairs in the highly imbalanced datasets to get a 1:10 ratio of positive-to-negative samples during training.

3.2 Performance evaluation

Due to the high imbalance in the datasets, accuracy isn't a meaningful metric for our study. Instead, we relied on threshold-independent metrics AUROC and AUPRC which are suitable for imbalanced datasets [39]. We used the median AUROC of test complexes for evaluation in order to avoid extreme changes in score due to complex sizes [25].

3.2.1 Pairwise PPIS prediction results

Results on DBD5: For pairwise interaction site prediction, we compared the Median AUROC and AUPRC scores of Pair-EGRET with machine learning, CNN and GNN-based methods [11, 40–42, 10, 43–45]. It is evident from Table 1 that Pair-EGRET outperforms all the other methods with respect to our primary evaluation metric, median AUROC with a score of 0.88828 while achieving an AUPRC score of 0.0173 which is comparable to the best-performing (0.018) method DCNN[41].

Results on Dockground: Although the unbound benchmark version 4 of Dockground is relatively less explored for the pairwise PPIS prediction task, we evaluated Pair-EGRET on this benchmark because of the varying levels of difficulty it offers making it more diverse than DBD5. Compared to BIPSPI+ [46] – one of the few methods evaluated on Dockground for this task, Pair-EGRET performs very well with a median AUROC of 0.8747, highlighting its robustness in identifying interaction sites in relatively difficult complexes.

3.2.2 Interface region prediction results

Results on DBD5: We compared Pair-EGRET with BIPSI, BIPSPI+, and PInet [30] for interface region prediction on the DBD5 test dataset. In addition to AUROC and AUPRC, we also report the precision and recall scores of the methods for a fair comparison and consistency with the other baselines. Pair-EGRET outperforms all the other methods under two evaluation metrics - AUROC and recall. Remarkably, the AUROC score of Pair-EGRET is 0.924 which is 8.96% higher than the second best method BIPSPI+.

Table 1: Comparison between the predictive performance of different methods in predicting pairwise PPIS of the test complexes of DBD5 and Dockground. Scores for the baselines on DBD5 are directly reported from [10, 25, 26].

DBD5						Dockground	
Method	Median AUROC	AUPRC	Method	Median AUROC	AUPRC	Method	Median AUROC
BIPSPI	0.878	-	DTNN	0.867	0.007	BIPSPI+	0.831
SASNet	0.876	-	NEA	0.876	0.012	Pair-EGRET	**0.8747**
DCNN	0.828	**0.018**	EGNN	0.829	-		
NGF	0.865	0.007	GVP-GNN	0.885	-		
Pair-EGRET	**0.888**	0.0173					

Results on MaSIF: Our assessment of Pair-EGRET on the MaSIF dataset involved a comparison with MaSIF [9], SPPIDER [47], and PInet [30]. Pair-EGRET's superior performance is evident from its AUROC score of 0.9583, which is 8.89% than the next best method PInet. Additionally, Pair-EGRET outperforms all other models in terms of AUPRC (0.5938) as well. It's worth noting that, the use of bound conformations in the MaSIF dataset generally contributes to better performance compared to other benchmarks for all methods.

Table 2: Performance comparison of different methods in identifying interface regions of test complexes of DBD5 and a subset of MaSIF curated by the authors of [30]. geom* indicates models that only use geometric features of proteins.

DBD5					MaSIF		
Method	AUROC	AUPRC	Precision	Recall	Method	AUROC	AUPRC
BIPSPI	0.822	0.410	0.391	0.558	SPPDER	0.65	-
					MaSIF geom*	0.68	-
BIPSPI+	0.848	0.4653	0.438	0.573	MaSIF	0.87	-
					PInet geom*	0.75	0.30
PInet	0.753	**0.596**	**0.492**	0.723	PInet	0.88	0.45
Pair-EGRET	**0.924**	0.275	0.255	**0.746**	Pair-EGRET	**0.9583**	**0.5938**

3.3 Case study

We performed a visual comparison between the interface regions predicted by Pair-EGRET and a competitive method NEA (pairwise PPIS prediction method developed by Fout et al. [10]) for two representative complexes (PDB ID 3HI6 and 1JTD from DBD5 test set) using PyMOL software [48]. Comparing the results from Fig. B3, Pair-EGRET's predictions (green) were more concentrated around the true interface regions (yellow) and covered almost the entire interface. NEA's predictions (cyan) were more scattered and missed parts of the actual interface in both complexes.

4 Conclusions

In this study, we introduced Pair-EGRET, a novel deep-learning method for accurate pairwise interaction site and interface region prediction in protein complexes using an edge-aggregated graph attention network and cross-attention mechanism. Additionally, we explored extensions to the original EGRET architecture, such as incorporating physicochemical features, positional encoders, and multi-headed cross-attention layers. Future directions for this study include leveraging larger protein language models like ESM-2 [49] by Facebook Research and benefiting from the availability of larger datasets[40, 9] assembled from structure-known proteins for enhancing our model's performance.

References

[1] Javier De Las Rivas and Celia Fontanillo. Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS computational biology*, 6(6):e1000807, 2010.

[2] Uros Kuzmanov and Andrew Emili. Protein-protein interaction networks: probing disease mechanisms using model systems. *Genome medicine*, 5(4):1–12, 2013.

[3] Rod K Nibbe, Salim A Chowdhury, Mehmet Koyutürk, Rob Ewing, and Mark R Chance. Protein–protein interaction networks and subnetworks in the biology of disease. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 3(3):357–367, 2011.

[4] Ioanna Petta, Sam Lievens, Claude Libert, Jan Tavernier, and Karolien De Bosscher. Modulation of protein–protein interactions for the development of novel therapeutics. *Molecular Therapy*, 24(4):707–718, 2016.

[5] Olivier Sperandio. Toward the design of drugs on protein-protein interactions. *Current pharmaceutical design*, 18(30):4585–4585, 2012.

[6] Khaled S Ahmed, Nahed H Saloma, and Yasser M Kadah. Improving the prediction of yeast protein function using weighted protein-protein interactions. *Theoretical Biology and Medical Modelling*, 8:1–17, 2011.

[7] Naoki Orii and Madhavi K Ganapathiraju. Wiki-pi: a web-server of annotated human protein-protein interactions to aid in discovery of protein function. *PloS one*, 7(11):e49029, 2012.

[8] Sazan Mahbub and Md Shamsuzzoha Bayzid. Egret: edge aggregated graph attention networks and transfer learning improve protein–protein interaction site prediction. *Briefings in Bioinformatics*, 23(2):bbab578, 2022.

[9] Pablo Gainza, Freyr Sverrisson, F. Monti, Emanuele Rodolà, D. Boscaini, Michael M. Bronstein, and Bruno E. Correia. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, 17:184–192, 2019.

[10] Alex Fout, Jonathon Byrd, Basir Shariat, and Asa Ben-Hur. Protein interface prediction using graph convolutional networks. *Advances in neural information processing systems*, 30, 2017.

[11] Ruben Sanchez-Garcia, Carlos Oscar Sánchez Sorzano, José María Carazo, and Joan Segura. Bipspi: a method for the prediction of partner-specific protein–protein interfaces. *Bioinformatics*, 35(3):470–477, 2019.

[12] Stephen R Comeau, David W Gatchell, Sandor Vajda, and Carlos J Camacho. Cluspro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, 20(1):45–50, 2004.

[13] Brian G Pierce, Kevin Wiehe, Howook Hwang, Bong-Hyun Kim, Thom Vreven, and Zhiping Weng. Zdock server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics*, 30(12):1771–1773, 2014.

[14] Dina Schneidman-Duhovny, Yuval Inbar, Ruth Nussinov, and Haim J Wolfson. Patchdock and symmdock: servers for rigid and symmetric docking. *Nucleic acids research*, 33(suppl_2):W363–W367, 2005.

[15] Yumeng Yan, Huanyu Tao, Jiahua He, and Sheng-You Huang. The hdock server for integrated protein–protein docking. *Nature protocols*, 15(5):1829–1852, 2020.

[16] Roberto Mosca, Arnaud Céol, and Patrick Aloy. Interactome3d: adding structural details to protein networks. *Nature methods*, 10(1):47–53, 2013.

[17] Joan Segura, Ruben Sanchez-Garcia, Daniel Tabas-Madrid, Jesus Cuenca-Alba, Carlos Oscar S Sorzano, and Jose Maria Carazo. 3diana: 3d domain interaction analysis: a toolbox for quaternary structure modeling. *Biophysical Journal*, 110(4):766–775, 2016.

[18] Li C Xue, Drena Dobbs, and Vasant Honavar. Homppi: a class of sequence homology based protein-protein interface prediction methods. *BMC bioinformatics*, 12(1):1–24, 2011.

[19] James R Bradford and David R Westhead. Improved prediction of protein–protein binding sites using a support vector machines approach. *Bioinformatics*, 21(8):1487–1494, 2005.

[20] Mile Šikić, Sanja Tomić, and Kristian Vlahoviček. Prediction of protein–protein interaction sites in sequences and 3d structures by random forests. *PLoS computational biology*, 5(1):e1000278, 2009.

[21] Feihong Wu, Byron Olson, Drena Dobbs, and Vasant Honavar. Comparing kernels for predicting protein binding sites from amino acid sequence. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pages 1612–1616. IEEE, 2006.

[22] Thomas C Northey, Anja Barešić, and Andrew CR Martin. Intpred: a structure-based predictor of protein–protein interaction sites. *Bioinformatics*, 34(2):223–229, 2018.

[23] Xiaoying Wang, Bin Yu, Anjun Ma, Cheng Chen, Bingqiang Liu, and Qin Ma. Protein–protein interaction sites prediction by ensemble random forests with synthetic minority oversampling technique. *Bioinformatics*, 35(14):2395–2402, 2019.

[24] Qingzhen Hou, Paul FG De Geest, Wim F Vranken, Jaap Heringa, and K Anton Feenstra. Seeing the trees through the forest: sequence-based homo-and heteromeric protein-protein interaction sites prediction using random forest. *Bioinformatics*, 33(10):1479–1487, 2017.

[25] Yi Liu, Hao Yuan, Lei Cai, and Shuiwang Ji. Deep learning of high-order interactions for protein interface prediction. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 679–687, 2020.

[26] Fang Wu, Tao Yu, Dragomir Radev, and Jinbo Xu. When geometric deep learning meets pretrained protein language models. *CoRR*, abs/2212.03447, 2022.

[27] Jesse Vig, Ali Madani, Lav R Varshney, Caiming Xiong, Richard Socher, and Nazneen Fatema Rajani. Bertology meets biology: interpreting attention in protein language models. 2021.

[28] Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, et al. Prottrans: Toward understanding the language of life through self-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(10):7112–7127, 2021.

[29] Jérôme Tubiana, Dina Schneidman-Duhovny, and Haim J Wolfson. Scannet: an interpretable geometric deep learning model for structure-based protein binding site prediction. *Nature Methods*, 19(6):730–739, 2022.

[30] Bowen Dai and Chris Bailey-Kellogg. Protein interaction interface region prediction by geometric deep learning. *Bioinformatics*, 37(17):2580–2588, 2021.

[31] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.

[32] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. 2018.

[33] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[34] David Eppstein, Michael S Paterson, and F Frances Yao. On nearest-neighbor graphs. *Discrete & Computational Geometry*, 17:263–282, 1997.

[35] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

[36] Thom Vreven, Iain H Moal, Anna Vangone, Brian G Pierce, Panagiotis L Kastritis, Mieczyslaw Torchala, Raphael Chaleil, Brian Jiménez-García, Paul A Bates, Juan Fernandez-Recio, et al. Updates to the integrated protein–protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2. *Journal of molecular biology*, 427(19):3031–3041, 2015.

[37] Petras J Kundrotas, Ivan Anishchenko, Taras Dauzhenka, Ian Kotthoff, Daniil Mnevets, Matthew M Copeland, and Ilya A Vakser. Dockground: a comprehensive data resource for modeling of protein complexes. *Protein Science*, 27(1):172–181, 2018.

[38] Fayyaz ul Amir Afsar Minhas, Brian J Geiss, and Asa Ben-Hur. Pairpred: partner-specific prediction of interacting residues from sequence and structure. *Proteins: Structure, Function, and Bioinformatics*, 82(7): 1142–1155, 2014.

[39] Yiwei Li, G Brian Golding, and Lucian Ilie. Delphi: accurate deep ensemble model for protein interaction sites prediction. *Bioinformatics*, 37(7):896–904, 2021.

[40] Raphael Townshend, Rishi Bedi, Patricia Suriana, and Ron Dror. End-to-end learning on 3d protein structure for interface prediction. *Advances in Neural Information Processing Systems*, 32, 2019.

[41] James Atwood and Don Towsley. Diffusion-convolutional neural networks. *Advances in neural information processing systems*, 29, 2016.

[42] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. *Advances in neural information processing systems*, 28, 2015.

[43] Kristof T Schütt, Farhad Arbabzadah, Stefan Chmiela, Klaus R Müller, and Alexandre Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature communications*, 8(1):13890, 2017.

[44] Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pages 9323–9332. PMLR, 2021.

[45] Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael John Lamarre Townshend, and Ron Dror. Learning from protein structure with geometric vector perceptrons. In *International Conference on Learning Representations*, 2021. URL `https://openreview.net/forum?id=1YLJDvSx6J4`.

[46] R Sanchez-Garcia, JR Macias, COS Sorzano, JM Carazo, and J Segura. Bipspi+: Mining type-specific datasets of protein complexes to improve protein binding site prediction. *Journal of Molecular Biology*, 434(11):167556, 2022.

[47] Aleksey Porollo and Jarosław Meller. Prediction-based fingerprints of protein–protein interactions. *Proteins: Structure, Function, and Bioinformatics*, 66(3):630–645, 2007.

[48] Warren L DeLano. The pymol molecular graphics system. *http://www. pymol. org/*, 2002.

[49] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Sal Candido, et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *BioRxiv*, 2022:500902, 2022.

[50] Soumyadeep Debnath and Ayatullah Faruk Mollah. A supervised machine learning approach for sequence based protein-protein interaction (ppi) prediction. *CoRR*, abs/2203.12659, 2022.

[51] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[52] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *CoRR*, abs/1607.06450, 2016.

[53] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28(1):235–242, 2000.

[54] Howook Hwang, Thom Vreven, Joël Janin, and Zhiping Weng. Protein–protein docking benchmark version 4.0. *Proteins: Structure, Function, and Bioinformatics*, 78(15):3111–3114, 2010.

[55] Utkan Ogmen, Ozlem Keskin, A Selim Aytuna, Ruth Nussinov, and Attila Gursoy. Prism: protein interactions by structural matching. *Nucleic acids research*, 33(suppl_2):W331–W336, 2005.

[56] Renxiao Wang, Xueliang Fang, Yipin Lu, Chao-Yie Yang, and Shaomeng Wang. The pdbbind database: methodologies and updates. *Journal of medicinal chemistry*, 48(12):4111–4119, 2005.

[57] James Dunbar, Konrad Krawczyk, Jinwoo Leem, Terry Baker, Angelika Fuchs, Guy Georges, Jiye Shi, and Charlotte M Deane. Sabdab: the structural antibody database. *Nucleic acids research*, 42(D1): D1140–D1146, 2014.

Appendix A: Model details

A.1 Features of protein graph

Node-level features

Each residue i in a protein is associated with a feature vector $q_i \in \mathbb{R}^{d_{node}}$, where d_{node} represents the number of node features used in this study. We leverage two types of node features to represent a residue.

i) Embedding-based features of the residues were extracted from the protein sequences using ProtBERT, a contextual embedding generation pipeline developed by [27]. ProtBERT captures both local and global context, including neighboring residues and overall protein structure, to generate embedding vectors $e = \{e_1, e_2, ..., e_N\}$, $e_i \in \mathbb{R}^{d_{protbert}}$ ($d_{protbert} = 1024$ and N = number of residues in the protein), which encode the structural and functional characteristics of the residues. Alternatively, other embedding generation models available in ProtTrans [28], such as ProtXL or ProtXLNet, can be used instead of ProtBERT without significantly impacting performance.

ii) Physicochemical Features of amino acids were incorporated as node features. These features encompass a range of properties, including hydrophilicity, flexibility, accessibility, turns scale, exposed surface, polarity, antigenic propensity, hydrophobicity, net charge index of side chains, polarizability, solvent-accessible surface area (SASA), relative SASA, side-chain volume, and residue depth. Notably, we calculate relative hydrophobicity and polarity based on two different scales or methods, namely H11a, H12a, P11a, and P12a, respectively, to ensure a comprehensive representation of these characteristics in our analysis [50]. These properties of a node i is represented by a vector $p_i \in \mathbb{R}^{d_{phychem}}$ ($d_{phychem} = 16$).
By concatenating the vectors e_i and p_i we obtained the final node features $q = \{q_1, q_2, ..., q_N\}$, $q_i \in \mathbb{R}^{d_{node}}$ where $d_{node} = d_{protbert} + d_{phychem} = 1024 + 16 = 1040$.

Edge-level features

Similar to EGRET [8], we utilize two edge-level features to represent the relationship between residues i and j through the edge ξ_{ij} where $\xi_{ij} \in \mathbb{R}^{f_e}$ and $f_e = 2$: (i) Inter-residue distance D_{ij} which denotes the average distance between the atoms of the residues and (ii) Relative orientation θ_{ij} which is measured by the absolute value of the angle formed by the surface-normals of the planes passing through the alpha carbon atom (C_α), the carbon atom of the carboxyl group, and the nitrogen atom of the amino group of each residue.

A.2 Components of EGRET

We discuss the structures and functionalities of the three core components of the EGRET model to make this paper self-contained and comprehensive.

i) Local feature extractor

The local feature extractor captures local interactions of the protein residues with other "sequentially closer" residues (not necessarily close in Euclidean space) while reducing the dimensionality of the node-level features. A one-dimensional convolutional neural network with a small odd number window size is used to encode the node feature vectors $q = \{q_1, q_2, ..., q_N\}$ into a new condensed and neighbor-aware feature representation $h = \{h_1, h_2, ..., h_N\}$, $h_i \in \mathbb{R}^{f_n}$, where $f_n < d_{protbert}$.

ii) Edge-aggregated graph attention layer

The edge-aggregated graph attention layer transforms the features h_i of the node i by encoding the three-dimensional structural information of its neighborhood N_i. This layer uses a modified version of the original graph attention layer [32] and aggregation process used in various GNN-based architectures [35, 32]. In the original aggregation process, the node features are transformed by taking a weighted average of the neighborhood node features using the equation: $\hat{h_i} = \sigma(\sum_{j \in N_i} \gamma_{ij} W^v h_j)$ where, $W^v \in \mathbb{R}^{f_n \times f_n}$ is a learnable parameter and γ_{ij} is the attention score calculated from h_i and

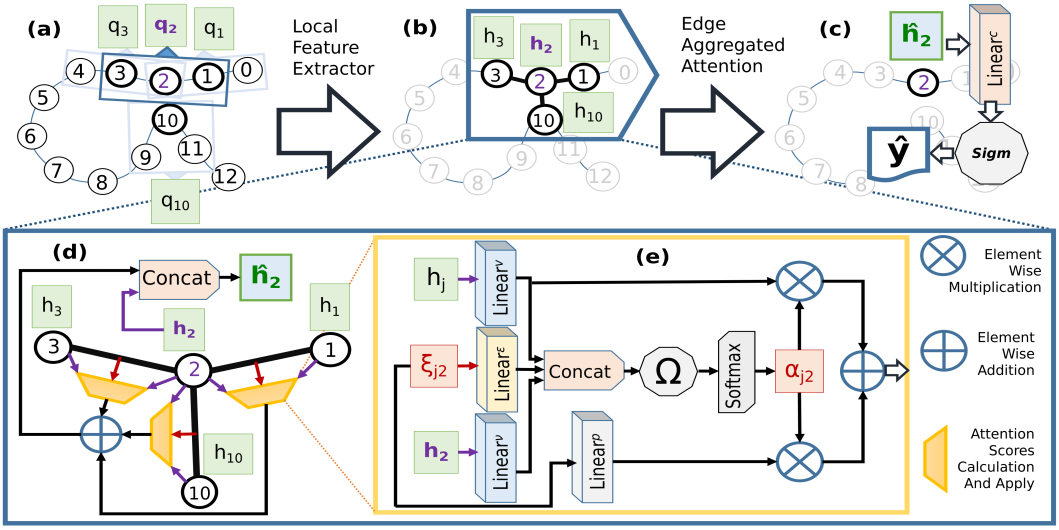


Fig. A1: Schematic diagram of the overall pipeline of EGRET being applied to a dummy protein having 13 residues. **(a)** Local feature extractor (with window size w_{local} = 3). **(b)** Edge-aggregated graph attention layer applied to residue 2 with neighborhood $N_2 = \{1, 3, 10\}$. **(c)** Node level classifier applied to final representation $\hat{h_2}$ of node 2. **(d)** The details of the edge-aggregated graph attention layer in an expanded form. **(e)** The expanded form of the module that calculates the attention scores for aggregation. [All figures were taken from [8]]

h_j that represents the importance of the features of node j to node i. EGRET improves upon this method by incorporating edge features during the calculation of attention scores and the aggregation process, resulting in a new scoring function e_{ji} and attention distribution α_{ji}. These metrics are obtained from the following equations.

$$e_{ji} = \Omega(W^{\alpha}[W^{v}h_i||W^{v}h_j||W^{p}\xi_{ji}]) \tag{1}$$

$$\alpha_{ji} = softmax(e_{ji}) = \frac{exp(e_{ji})}{\sum_{k\in N_i} exp(e_{ki})} \tag{2}$$

Finally, the node and edge features are aggregated using the equation

$$\hat{h_i} = \sigma(\sum_{j\in N_i} \alpha_{ji}W^{v}h_j + \sum_{j\in N_i} \alpha_{ji}W^{\epsilon}\xi_{ji}) \,||\, h_i \tag{3}$$

Here, $W^{\alpha} \in \mathbb{R}^{2f_n+f_e}$, $W^{v} \in \mathbb{R}^{f_n\times f_n}$, $W^{p} \in \mathbb{R}^{f_e\times f_e}$ and $W^{\epsilon} \in \mathbb{R}^{f_n\times f_e}$ are learnable parameters, $||$ is the concatenation operator, and $\Omega(.)$ and $\sigma(.)$ are activation functions.

iii) Node-level classifier

The node-level classifier of EGRET linearly transforms the output of the previous layer $\hat{h_i}$. It applies the sigmoid activation function to generate the probability of being an interaction site for the residue represented by the node i.

A.3 Components of Pair-EGRET

The first three components of Pair-EGRET (i.e. Siamese EGRET network, positional encoder, and multi-headed cross-attention layer) are connected sequentially and are common to the architecture required for both pairwise interaction site prediction and interface region identification problems. The final two layers (pairwise classifier and interface region classifier) are parallel networks that generate the outputs corresponding to these two problems.

i) Siamese EGRET network

Each identical branch of the Siamese EGRET network contains only the first two modules of EGRET. Since the third module of EGRET (the node-level classifier) produces the probability of interaction for each residue without any knowledge of the partner protein in the complex, we discard this layer and instead utilize the features obtained from the local feature extractor and GAT layer of EGRET to encode each residue with features derived from its neighborhood, taking into account both sequential and spatial proximity.

The node features q_i and edge features ξ_{ij} corresponding to each of the graphs $G_{receptor}$ and G_{ligand} are fed through the siamese network to generate features $\hat{h}^l$ and $\hat{h}^r$ respectively. In this study, along with the embedding-based features used in the original EGRET paper, we also incorporate a wide range of physicochemical features for each residue to use as node features. Additionally, we enhance the quality of the features generated from the Siamese network by increasing the number of convolution layers in the local feature extractor and using multiple layers of the graph attention module.

Weight sharing between the branches of the Siamese network ensures that the model learns a common representation of proteins and is invariant to the ordering of the partner proteins provided to it.

ii) Positional encoder

According to Vaswani et al. [33], positional encoding for position pos in the sequence can be defined as:

$$PE(pos, i) = \begin{cases} sin(pos/10000^{\frac{i}{d_{model}}}), & \text{if } i \text{ is even,} \\ cos(pos/10000^{\frac{i-1}{d_{model}}}), & \text{otherwise,} \end{cases} \tag{4}$$

where $i \in [1, d_{model}]$ is the dimension of the embedding vector being calculated, and d_{model} is the dimension of the input embedding vector. This function maps the position of each amino acid to a set of position-specific embeddings PE. PE is added to the feature representations $\hat{h}^l$ and $\hat{h}^r$ obtained from the siamese network, to generate the outputs P^l and P^r of this layer.

$$P^l_{pos} = \hat{h}^l_{pos} + PE_{pos} \ , \ P^r_{pos} = \hat{h}^r_{pos} + PE_{pos}$$

iii) Multi-headed cross-attention layer

The structure of the multi-headed cross-attention layer is inspired by the encoder-decoder attention module described by [33]. The cross-attention module transforms the input feature vectors into three feature spaces: query, key, and value using learnable parameters W_Q, W_K, and $W_V \in \mathbb{R}^{d_k \times d_k}$, where d_k is a hyperparameter. Specifically, the ligand feature vector P^l obtained from the positional encoder layer is transformed into query $Q^l = W_Q P^l$, key $K^l = W_K P^l$, and value $V^l = W_V P^l$, while the receptor feature vector P^r is transformed into Q^r, K^r, and V^r using similar equations.

To model the interaction between the proteins, the cross-attention module is applied to the ligand and receptor feature vectors separately, as defined by the following equations:

$$\begin{aligned} Attention(Q^l, K^r, V^r) &= softmax(\frac{Q^l (K^r)^T}{\sqrt{d_k}}) V^r \\ Attention(Q^r, K^l, V^l) &= softmax(\frac{Q^r (K^l)^T}{\sqrt{d_k}}) V^l \end{aligned} \tag{5}$$

where $softmax$ is the softmax activation function, and $(\cdot)^T$ denotes the transpose operation. Each attention head in the multi-headed cross-attention module is an independent attention layer that captures different aspects of the input. Specifically, multi-headed attention can be described by the equation:

$$MultiHead(Q, K, V) = W_O Concat([head_1, head_2, \ldots, head_{N_h}])$$

where $head_i = Attention(Q, K, V)$ is the output of the i-th attention head , N_h is the number of total attention heads, and W_O is a learnable parameter used to project the concatenated attention heads back to the original feature space.

The outputs of the multi-headed attention are then added to the input of the module through a residual connection [51] which enables the model to capture both the attended features and original features,

followed by layer normalization [52] which stabilizes the model. The final outputs M^l and M^r are defined by the following equations:

$$M^l = LayerNorm(P^l + MultiHead(Q^l, K^r, V^r))$$
$$M^r = LayerNorm(P^r + MultiHead(Q^r, K^l, V^l))$$

iv) Pairwise classifier

For the task of pairwise interaction site prediction, for each residue pair (l_i, r_i) we have to generate an output $O_i \in \{0, 1\}$ where $O_i = 1$ would indicate that residue number l_i of the ligand and residue number r_i of the receptor are an interacting pair, whereas $O_i = 0$ would indicate otherwise.

In the pairwise classifier layer, for each residue pair, we extract the node features $M^l_{l_i}$ and $M^r_{r_i}$ from the outputs M^l and M^l generated by the cross-attention module and concatenate them. This representation of the residue pair is passed through a feed-forward neural network that incorporates a combination of learned weights, nonlinear activation functions such as ReLU and LeakyReLU, and finally, a sigmoid activation function to predict the probability of interaction between the nodes l_i and r_i.

Similar to [10], to ensure invariance to the ordering of the receptor and ligand, we concatenate the features in both possible orders, resulting in two predictions:

$$O^{rl}_i = \sigma(\text{FFN}(M^r r_i \| M^l l_i)) \quad , \quad O^{lr}_i = \sigma(\text{FFN}(M^l l_i \| M^r r_i))$$

where FFN is the feed forward network and σ is the Sigmoid activation function. Finally, we take the average of these two predictions to generate the final output probability O_i for the residue pair (l_i, r_i):

$$O_i = \frac{1}{2}(O^{rl}_i + O^{lr}_i)$$

v) Interface Region Classifier

For the task of interface region prediction, for each residue l_i of the ligand or r_j of the receptor we have to generate outputs O^l_i or O^r_j which would indicate whether the residue is a part of the interface of the complex or not. In the interface region classifier layer, instead of concatenating the outputs M^l and M^r generated by the cross-attention layer, we pass M^l and M^r individually through a feed-forward network. The dense layers linearly transform the feature vectors and the sigmoid activation function is applied to generate outputs O^l and O^r for the ligand and receptor proteins respectively.

$$O^l_i = \text{Dense}(M^l l_i) \quad , \quad O^r_j = \text{Dense}(M^r r_j)$$

The outputs represent the probabilities of each residue of a protein being part of the interface region of the complex.

Appendix B: Experiment details

B.1 Summary of datasets

i) Docking Benchmark version 5.0 (DBD5) is widely recognized as the standard benchmark for evaluating pairwise PPIS prediction and interface region identification. The dataset includes structures of 230 complexes from the protein data bank (PDB) [53] with amino acid sequence lengths of the constituent proteins varying from 29 to 2128. Training and validation on DBD5 were performed using the 175 complexes present in version 4.0 of Docking Benchmark (DBD4) [54]. We performed an 80%-20% partition of the 175 complexes stratifying them by the difficulty provided in [36]. For testing, we used a set of 55 complexes that were added in the update from DBD4 to DBD5. This time-based split of the dataset simulates the ability of the model to predict unreleased complexes, as opposed to a random split which has more training/testing cross-contamination. [40]

ii) Dockground is another benchmark used for evaluating pairwise interaction site prediction models. The dataset contains a diverse array of protein complexes of varying difficulties. Compared to DBD5, it has relatively fewer proteins with rigid bodies and more with higher difficulty levels. In

Table B1: Summary of the datasets used in this study

Dataset	Samples	Train	Validation	Test	Total
DBD5	Complexes	140	35	55	230
	Positive samples	12,866 (9.09%)	3,138 (0.2%)	4,871 (0.1%)	20,875 (0.3%)
	Negative samples	128,660 (90.9%)	1,874,322 (99.8%)	4,953,446 (99.9%)	6,956,428 (99.7%)
Dockground	Complexes	236	60	100	396
	Positive samples	14,007 (9.09%)	3,940 (0.05%)	5,905 (0.04%)	23,852 (0.11%)
	Negative samples	140,070 (90.9%)	7,199,540 (99.94%)	12,673,885 (99.95%)	20,013,495 (99.88%)
MaSIF	Complexes	1890	470	787	3147
	Positive samples	308,441 (9.09 %)	77,903 (0.28%)	94,135 (0.26%)	480,479 (0.74%)
	Negative samples	3,084,229 (90.9%)	27,074,411 (99.71%)	34,833,241 (99.73%)	64,991,881 (99.26%)

our experiments, we used the unbound docking benchmark set 4 of the Dockground dataset containing 396 complexes with only 77 complexes shared with DBD5. We used 236 complexes for training, 60 complexes for validation, and 100 complexes for testing Pair-EGRET on the task of pairwise interaction site prediction.

(iii) MaSIF is a relatively large dataset containing a total of 3362 complexes taken from the PRISM [55] list of nonredundant proteins, the ZDock benchmark [13], PDBBind [56], and SabDab [57] dataset. For evaluating our model, we used the curated subset of MaSIF used by PInet [30] which excludes complexes with interface regions smaller than 1% of the size of the ligand. We also excluded complexes with receptor or ligand sequence lengths smaller than the minimum neighborhood size required by the edge-aggregated graph attention layer of EGRET. This resulted in a dataset containing 3147 complexes, which was split into 1890 training, 470 validation and, 787 test complexes. We used this subset of MaSIF for evaluating Pair-EGRET in identifying interface regions of complexes. This benchmark uses bound conformations of proteins to produce features for training the models. Consistent with other methods, only for this benchmark, we used the provided bound conformations for generating the node and edge-level features of Pair-EGRET.

B.2 Analysis of model performance

We analyzed the impact of different features and modules of Pair-EGRET on its performance and visualized the patterns of cross-attention scores generated by the model to improve the interpretability of the results found in this study.

Impact of different node-level features: Table B2 shows the impact of different node-level features on the median AUROC scores of Pair-EGRET in predicting pairwise PPIS from DBD5 test complexes. The results highlight that adding ProtBERT-based and physicochemical features improves the median AUROC score of Pair-EGRET by 10.168% and 8.518% respectively, indicating that the model may be benefiting from the patterns captured by ProtBERT embeddings and the physical characteristics represented by the physicochemical features.

Impact of different modules of Pair-EGRET In Table B3, we analyzed the impact of different core modules of Pair-EGRET, particularly the positional encoder and the multi-headed cross-attention layer on its performance. The addition of the positional encoder introduces a 3.837% improvement in the median AUROC of Pair-EGRET for pairwise PPIS prediction in DBD5 complexes, while the cross-attention module introduces an improvement of 2.508%. This strengthens our argument that

the positional encoder enhances sequential context, and the cross-attention module enables residues to access relevant information from the residues of the partner protein.

We also conducted some ablation studies to assess the performance of models featuring simpler architectures when equipped with the enhanced physicochemical and ProtBERT-based features of Pair-EGRET. Specifically, we substituted the first three modules of Pair-EGRET with more straightforward layers and presented the pairwise PPIS prediction results for DBD5 test complexes using these modified models (see Table B4). The architectures we considered for this analysis include: a Siamese feed-forward network solely composed of fully connected layers, a CNN-based network employing 1D convolution operations on protein sequences, an attention-based model integrating a positional encoder and a cross-attention module, and a graph attention network [32] with a single graph convolution layer [35]. The results in the table reveal that the enhanced features used in this study yield reasonably good results even for very simple architectures such as Siamese FFN and 1D CNN. However, the incorporation of GAT or attention modules significantly improves the model's performance. Notably, the GAT network with a single convolution layer achieves the highest median AUROC score among these models. These findings suggest that the performance boost in Pair-EGRET can be attributed to both the enhanced features utilized in this study and the effective architecture of Pair-EGRET. The model not only relies on improved features but also excels in accurately capturing the contextual intricacies conveyed by these features, ultimately contributing to a significant overall performance improvement.

Table B2: Median AUROC scores of Pair-EGRET for predicting pairwise interaction sites in DBD5 test complexes, evaluated with various combinations of node-level features.

Node features combination	Median AUROC (DBD5)	Improvement by feature
With both features	**0.88828**	-
Without physicochemical features	0.8031	+8.518%
Without ProtBERT-based features	0.7866	+10.168%

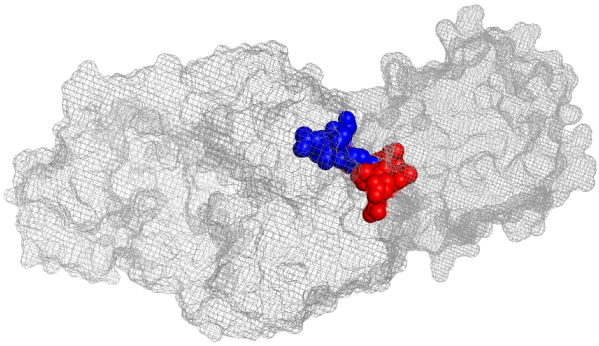
Table B3: Median AUROC scores of Pair-EGRET for predicting pairwise interaction sites in DBD5 test complexes, comparing the impact of different modules in the Pair-EGRET architecture.

Model architecture	Median AUROC (DBD5)	Improvement by module
Our full framework	**0.88828**	-
Without cross-attention	0.8632	+2.508%
Without positional encoder	0.84991	+3.837%

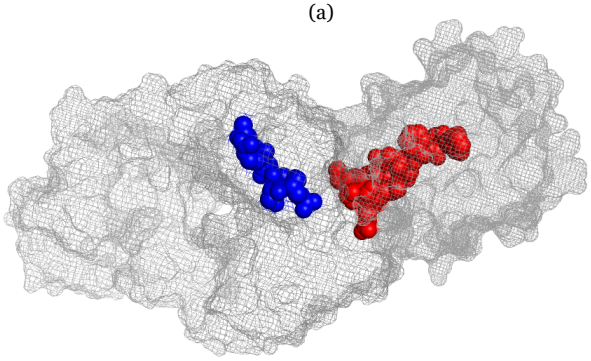
Table B4: Performance of models with simpler architectures compared to Pair-EGRET on PPIS prediction of DBD5 test complexes using the same set of enhanced features as Pair-EGRET.

Model architecture	Median AUROC (DBD5)
Siamese feed-forward neural network	0.6947
1D Convolutional neural network	0.7112
Attention-based network	0.7889
GAT with a single graph convolution layer	0.8232

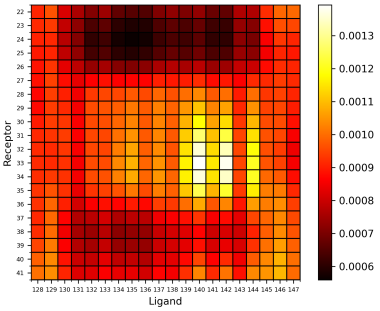
Patterns of cross-attention scores In Fig. B1(c), we present a heatmap of attention scores produced by the cross-attention layer for a representative protein (PDB ID 3HI6). Remarkably, the highest attention scores (lighter colors) in the heatmap correspond to interacting residue pairs or closely located pairs, as confirmed by PyMOL visualization Fig. B1(a). Conversely, pairs with lower attention scores (darker colors) are more distant and may even be buried within the proteins' surfaces Fig. B1(b). These findings highlight the meaningful connection between Pair-EGRET's cross-attention layer scores and the characteristics of interacting residue pairs.



(a)



(b)



(c)

Fig. B1: Patterns of cross-attention scores in a 20 residue window of chain A of the ligand and chain L of the receptor of a representative complex (PDB ID 3HI6). **(a)** PyMOL visualization of the residues corresponding to the lighter regions (high attention scores) of the heatmap. **(b)** PyMOL visualization of the residues corresponding to the darker regions (low attention scores) of the heatmap. **(c)** Heatmap of the attention scores generated by the multi-headed cross-attention layer of Pair-EGRET.

B.3 Additional figures

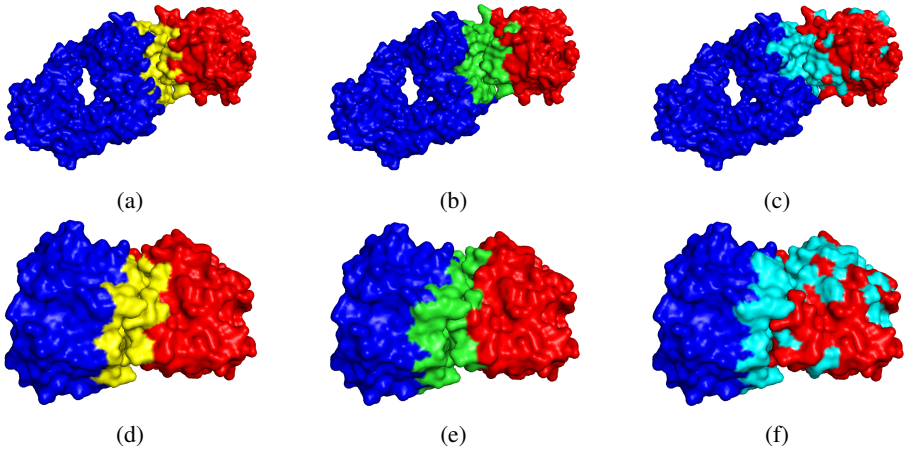


Fig. B2: PyMOL visualization of two representative complexes (PDB ID 3HI6 and 1JTD) from the test set of DBD5 in bound form. The red and blue surfaces represent the ligand and receptor proteins respectively. **(a-c)** The true interface regions of 3HI6 (yellow), the regions predicted by Pair-EGRET with 90-95% confidence (green), and the regions predicted by NEA with 75-80% confidence (cyan). **(d-f)** The true (yellow), predicted by Pair-EGRET (green), and predicted by NEA (cyan) interface regions of 1JTD.