Improved Graph-based Antibody-aware Epitope Prediction with Protein Language Model-based Embeddings

Mansoor Ahmed[1], Sarwan Ali[2], Avais Jan[1], Imdad Ullah Khan[3,*], and Murray Patterson[1,*]

[1] Georgia State University, Atlanta, GA, USA
{mahmed76,ajan3}@student.gsu.edu, mpatterson30@gsu.edu
[2] Columbia University Irving Medical Center, New York, NY, USA
sa4559@columbia.edu
[3] Lahore University of Management Sciences, Lahore, Pakistan
imdad.khan@lums.edu.pk * Corresponding authors, Equal contribution

Abstract. A major challenge in computational antibody design is the accurate identification of antigen binding site, i.e., epitope. The current approaches to epitope prediction struggle because of the variational nature of epitopes and the lack of availability of experimental datasets. However, deep learning-based approaches have shown great promise in achieving better results for the epitope prediction task in recent years. Moreover, there is now great potential in epitope-prediction research because of the newly released and largest-of-its-kind benchmark dataset, Antibody-specific Epitope Prediction (AsEP), modeling antibody-antigen complexes as graph pairs. In this paper, we employ a graph convolutional network (GCN) coupled with protein language models (PLM)-based residue embeddings for epitope prediction on the AsEP dataset. We explore the use of different PLM-embedding methods on the epitope prediction task and show that antibody-specific PLMs such as AntiBERTy and general PLMs such as ProtBERT and ESM2 for antigens provide improved epitope prediction performance with an area under the ROC curve of 0.65, precision of 0.28, and recall of 0.46. The source code is available at: https://github.com/mansoor181/walle-pp.git.

Keywords: Antibody design · Protein Language Models · Epitope prediction · Antigen-antibody interaction · Artificial Intelligence · Deep Learning · Protein Graphs

1 Introduction

Antibodies are large, Y-shaped proteins produced by B cells that play a critical role in the immune system by identifying and neutralizing foreign substances such as toxins, bacteria, and viruses, collectively known as antigens. Antibodies are known to be the largest class of biotherapeutics, where five of the current top 10 blockbuster drugs are monoclonal antibodies [1]. Hence, the study of antibodies is of significant importance in immunology and therapeutic development, as antibodies have wide applications in

diagnostics, therapeutics, and research [2,3]. However, traditional *in-vivo* methods for antibody production, such as immunizing animals, are time-consuming, expensive, and laborious [4,5]. These challenges make it imperative to develop computational methods for antibody design, which aim to streamline the process and reduce the antibody search space by producing cost-effective antibodies [6].

A critical step in antibody design is epitope prediction, which involves identifying the specific regions on the surface of an antigen that are recognized and bound by an antibody [7]. Accurate epitope prediction is essential for understanding antibody-antigen interactions and designing antibodies with high binding affinity [4]. However, this problem is inherently challenging due to several factors [8,9] . First, epitopes are often non-linear and conformationally diverse, making it difficult to capture their spatial relationships using sequence-based methods alone [5]. Second, a single antigen can have multiple epitopes, which further complicates the prediction task [1]. Finally, most of the current approaches rely on sequence-based data for epitope prediction, which fail to model the complex spatial arrangements of antigen binding sites [2].

Current approaches to epitope prediction can be broadly categorized into sequence-based and structure-based methods [10]. Sequence-based methods rely on amino acid sequences to predict epitopes, while structure-based methods leverage the 3D structure of antigens to identify potential binding sites [11]. Additionally, these methods can be classified as antibody-agnostic or antibody-aware [12]. Antibody-agnostic methods predict epitopes without considering the specific antibody, whereas antibody-aware methods incorporate information about the antibody's structure and binding interaction. The latter approaches have proved to be more accurate as they enable the identification of specific epitopes on the antigen surface that are recognized by a particular antibody [1]. Graph-based approaches, which represent antigens and antibodies as graphs, have shown further promise due to their ability to model complex spatial relationships [7,10] . However, the choice of embedding methods for representing residues in these graphs is critical for achieving accurate predictions [13].

In this work, we explore the use of protein language models (PLMs) for embedding residues in graph-based antibody-aware epitope prediction. Specifically, we evaluate the performance of state-of-the-art PLMs such as ESM2 [14], AntiBERTy [15], ProtBERT [16], and ESM-IF [17], alongside classical embedding methods like one-hot encoding and BLOSUM62. These embeddings are integrated into a graph convolutional network (GCN) framework to predict epitopes on antigen surfaces. Therefore, the antibody-aware epitope prediction problem can be posed as a binary classification problem: for each amino acid residue in the antibody and antigen, do they bind to each other or not? Our main contributions are as follows:

1. We provide a comprehensive comparison of different PLM-based embedding methods for the epitope prediction problem, highlighting their strengths and limitations.
2. We propose the use of AntiBERTy for antibody residue embedding and ProtBERT for antigen embedding, along with the GCN model for improved epitope prediction.

The remainder of this paper is organized as follows. Section 2 discusses related literature on epitope prediction and graph-based methods. Section 3 presents the proposed approach with the problem formulation, data representation, and model architecture.

Section 4 discusses the dataset and model training details. Section 5 discusses the experimental results and comparisons with baseline methods and outlines future research directions. Finally, Section 6 concludes the paper.

2 Related Work

Epitope prediction methods can be broadly categorized into antibody-agnostic and antibody-aware approaches. Antibody-agnostic methods predict epitopes without considering the specific antibody, focusing solely on the antigen's properties. These methods often rely on sequence-based features, structural information, or a combination of both. For instance, [3] proposed *epitope1D*, a taxonomy-aware method for linear B-cell epitope prediction using graph-based sequence representations. Similarly, [14] introduced *ESMFold*, a language model for protein structure prediction, which has been extended by [9] to predict protein binding sites using graph transformers. These methods, while effective for general epitope prediction, often lack the specificity required for antibody-specific applications. On the other hand, antibody-aware methods explicitly incorporate information about the antibody's structure or sequence to predict epitopes for a given antibody-antigen pair. For example, [8] developed *EpiScan*, an attention-based deep learning framework that uses antibody sequence information to predict antibody-specific epitopes. The authors in [18] proposed ESMBind where they finetuned complex ESM2 models ranging from 8M to 650 parameters such as `esm2_t33_650M_UR50D` on an annotated protein binding sites dataset, specifically for protein binding site prediction task. Similarly, [13] introduced *AsEP*, a benchmark dataset and a graph-based method (*WALLE*) that leverages protein language models and graph neural networks for antibody-specific epitope prediction. These methods demonstrate the importance of integrating antibody-specific information for accurate epitope prediction.

Recent advances in antibody-aware epitope prediction have highlighted the benefits of combining sequence and structural information. For instance, [10] proposed *EpiGraph*, which uses graph attention networks to capture the spatial clustering properties of epitopes, achieving state-of-the-art performance. Similarly, [7] introduced *GraphBepi*, a graph-based model that leverages AlphaFold2-predicted structures and ESM2 embeddings for B-cell epitope prediction. These methods underscore the importance of graph-based representations in capturing the spatial relationships between residues. Additionally, [19] developed *ClusPro AbEMap*, a docking-based server for antibody epitope mapping, which supports inputs ranging from X-ray structures to amino acid sequences. These approaches demonstrate the growing trend of integrating structural and sequence-based features to improve epitope prediction accuracy. However, challenges remain, particularly in generalizing to unseen epitopes and handling the limited availability of high-quality datasets.

Graph-based approaches have emerged as a powerful paradigm for epitope prediction, leveraging the inherent spatial relationships in protein structures. Methods like *PECAN* [20], *PInet* [21], and [22] use graph neural networks to predict protein-protein interaction interfaces, including antibody-antigen binding sites. These methods often employ transfer learning and attention mechanisms to improve performance and interpretability. For example, [11] proposed *EPMP*, a neural message-passing framework

that uses asymmetrical architectures for paratope and epitope prediction, achieving significant improvements over existing methods. Similarly, *EpiPred* [12] combines conformational matching and antibody-antigen scoring to improve epitope prediction and docking results. Moreover, the authors in [23] used GCN to capture information of spatial neighbors of a target residue while Attention-Based Bidirectional Long Short-Term Memory (Att-BLSTM) networks to extract global information from the whole antigen sequence. These graph-based methods highlight the potential of combining structural, sequential, and contextual information for accurate and interpretable epitope prediction. Despite these advances, there is a need for more comprehensive benchmarking and the development of methods that can handle the diversity and complexity of antibody-antigen interactions.

3 Proposed Approach

This section presents the problem formulation of the epitope prediction task, the data representation strategies employed, and the overall architecture of the proposed method.

3.1 Problem formulation

The interaction between antibodies and antigens can be framed as predicting links between the vertices of two disjoint undirected graphs: an antibody graph $G_A = (V_A, E_A)$ and an antigen graph $G_B = (V_B, E_B)$, i.e., $V_A \cap V_B = \emptyset$. Here, V_A and V_B represent the sets of vertices (residues) for the antibody and antigen graphs, respectively, while E_A and E_B represent their sets of edges based on residue proximity. Each vertex is encoded into a vector using a function $h_x : V_x \to \mathbb{R}^{D_x}$ where $x \in \{A, B\}$ and h can be any encoding function such as one-hot encoding or pre-trained embeddings from a PLM. The adjacency matrix $E_x \in \{0, 1\}^{|V_x| \times |V_x|}$ is derived from the distance matrix of the residues of antigen or antibody. Each entry e_{ij} indicates the proximity between residue i and residue j; $e_{ij} = 1$ if the Euclidean distance between any non-hydrogen atoms of residue i and residue j is less than 4.5Å, and $e_{ij} = 0$ otherwise. In order to make the computations faster, the antibody graph G_A is constructed only from the CDR residues of the antibody's heavy and light chains, while the antigen graph G_B is constructed from the surface residues of the antigen.

We address the following two key tasks in this framework:

Epitope Prediction This task involves identifying antigen residues that interact with the antibody. A residue $v \in V_B$ is labeled as an epitope (1) if it is within 4.5Å of any residue in V_A; otherwise, it is labeled as a non-epitope (0). Formally, the task is a binary node classification problem on the antigen graph G_B, conditioned on the antibody graph G_A. The classifier $f : V_B \to \{0, 1\}$ is defined as:

$$f(v; G_B, G_A) = \begin{cases} 1 & \text{if } v \text{ is an epitope,} \\ 0 & \text{otherwise.} \end{cases}$$

Bipartite Link Prediction This task predicts interactions between antibody and antigen residues, forming a bipartite graph $K_{m,n} = (V_A, V_B, E)$, where $m = |V_A|$, $n = |V_B|$, and E represents potential inter-graph edges. An edge $e \in E$ is labeled as 1 if the corresponding residues are within 4.5Å of each other, and 0 otherwise. The task is formulated as a binary edge classification problem, with the classifier $g : K_{m,n} \to \{0, 1\}$ defined as:

$$g(v_a, v_b; K_{m,n}) = \begin{cases} 1 & \text{if } v_a \text{ and } v_b \text{ are in contact,} \\ 0 & \text{otherwise.} \end{cases}$$

The problem formulation is derived partially from [10,13,20], please refer to these papers for details.

3.2 Graph Representation

Each antibody-antigen complex was converted into a graph representation, where protein residues are modeled as vertices. Edges are drawn between residues if their non-hydrogen atoms are within 4.5Å of each other. We focused on surface residues, excluding buried residues with zero solvent-accessible surface area from the antigen graph, and non-CDR residues from the antibody graph.

In our study, we employed various embedding methods to represent protein residues as continuous vectors. These methods include both classical techniques and state-of-the-art PLMs. Specifically, we employ AntiBERTy [4], an antibody-specific transformer language model pre-trained on 558M natural antibody sequences, which produces embeddings of fixed size $L \times 512$ where L represents the length of the sequence. Additionally, we use ESM2 [5], a general-purpose PLM that captures the evolutionary features of the protein by generating embeddings of size $L \times 1280$ using the `esm2_t33_650M_UR50D` model. Conversely, WALLE [13] used a smaller ESM2 model `esm2_t12_35M_UR50D` to produce embeddings of size $L \times 480$. ESM-IF1, an inverse folding model, is used to design sequences for given structures, producing embeddings of size $L \times 512$ with the `esm_if1_gvp4_t16_142M_UR50` model. ProtBERT [6], another advanced PLM, generates embeddings of size $L \times 1024$ with the model checkpoint `prot_t5_xl_uniref50`. For comparison, we also include classical embedding methods such as BLOSUM62 and One-Hot Encoding. These diverse embedding techniques allow us to capture both sequential and structural information effectively, enhancing the performance of our predictive models.

3.3 Model Architecture

The overall architecture of our graph-based framework for antibody-aware epitope prediction is shown in Figure 1. The model takes as input a pair of antibody and antigen graphs, constructed as detailed in Section 4.1, and performs node- and edge-level predictions.

[4] https://github.com/jeffreyruffolo/AntiBERTy.git
[5] https://github.com/facebookresearch/esm.git
[6] https://github.com/agemagician/ProtTrans.git

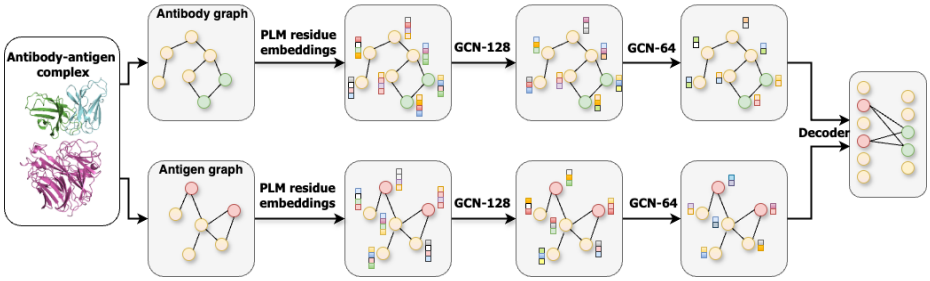


Fig. 1: The schematic of representing an antibody-antigen complex structure as a graph pair, employing protein language model (PLM)-based residue embeddings, and the graph convolutional network (GCN) model architecture. The yellow-colored nodes represent non-binding residues, while green and red-colored nodes represent paratopes and epitopes, respectively.

The architecture of our model incorporates graph modules that process the input graphs of antibody and antigen structures separately, inspired by WALLE, PECAN, and EPMP. The antibody graph is represented by node embeddings X_A with a shape of $m \times D_A$ and an adjacency matrix E_A, while the antigen graph is described by node embeddings X_B with a shape of $n \times D_B$ and its corresponding adjacency matrix E_B. The embedding sizes D_A and D_B vary depending on the embedding method used, such as AntiBERTy, ESM2, or other methods.

Both antibody and antigen graph nodes are first projected into a dimensionality of 128 using fully connected layers. The resulting embeddings are then passed through two Graph Convolutional Network (GCN) modules consecutively to refine the features and yield updated node embeddings X'_A and X'_B with a reduced dimensionality of $m \times 64$ and $n \times 64$, respectively. The output from the first GCN layer is passed through a ReLU activation function. Outputs from the second GCN layer are directly fed into the Decoder module. These GCNs operate independently, each with its own parameters, ensuring that the learned representations are specific to the antibody or the antigen.

The use of separate GCN modules for the antibody and antigen allows for the capture of unique structural and functional characteristics pertinent to each molecule before any interaction analysis. This design choice aligns with the understanding that the antibody and antigen have distinct roles in their interactions and their molecular features should be processed separately.

A simple decoder is employed to predict the binary labels of edges between the antibody and antigen graphs. The decoder takes a pair of node embeddings output by the graph modules, X'_A and X'_B, as input and constructs a bipartite adjacency matrix of size $m \times n$. Then, it predicts the probability of each edge: an edge is assigned a binary label of 1 if the predicted probability is greater than 0.5, or 0 otherwise. For the epitope prediction task, edge-level predictions are converted to node-level by summing the predicted probabilities of all edges connected to an antigen node. An antigen node is assigned a label of 1 if the number of connected edges is greater than a threshold,

or 0 otherwise. The threshold is treated as a hyperparameter and is optimized in the experiments.

4 Experimental Setup

In this section, we provide a description of the dataset, accompanied by an exploratory data analysis and dataset split strategies, and the implementation details of our proposed framework.

4.1 Dataset

We utilize the **AsEP** dataset [13], a novel benchmark dataset of antibody-antigen complexes designed specifically for epitope prediction tasks. The dataset is derived from the Antibody Database (AbDb) and the Protein Data Bank (PDB), comprising 1,723 unique antibody-antigen complexes after filtering and deduplication. Each complex includes a single-chain protein antigen paired with a conventional antibody containing both heavy and light chains. Non-canonical residues and unresolved CDR regions were excluded to ensure data quality.

Exploratory Data Analysis (EDA) Our EDA revealed several key insights into the dataset and are shown in Figure 2. The distribution of epitope residues showed a mean of 19 ± 4.7, while the antigen surface residues numbered in the hundreds. The contact distribution between residues in the bipartite graph had a mean of 43.7 contacts with a standard deviation of 12.8. Additionally, the dataset includes 641 unique antigens and 973 epitope groups, highlighting the diversity and complexity of the antibody-antigen interactions captured in the AsEP dataset.

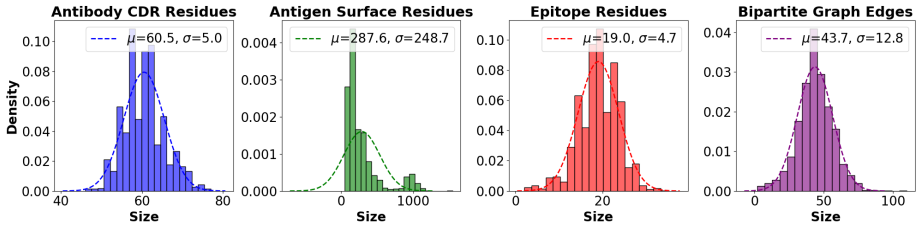


Fig. 2: (left to right) The distribution of the size of antibody CDR residues, antigen surface residues, epitope residues, and bipartite graph edges (antibody-antigen residue-residue contacts) in the dataset. The dotted lines represent the fitted normal distribution.

4.2 Implementation Details

The model implementation details were followed in part from WALLE [13]. We used the PyTorch Geometric framework to build the graph modules in the proposed model.

We used a weighted loss function consisting of a sum of binary cross-entropy loss for the bipartite graph link reconstruction and a regularizer for the number of positive edges in the reconstructed bipartite graph to train the model. We used the mean of edges in the antibody-antigen bipartite graph (as shown in Figure 2) to set the regularizer in the loss function to constrain the model not to predict too many or too few positive edges. We used two configurations for the decoder: a fully connected linear layer (GCN-L) and an inner product decoder (GCN). We used the same set of hyperparameters and loss functions for both dataset split settings. We evaluate the performance of our model using the standard classification metrics of Matthews correlation coefficient (MCC), the area under the precision-recall (AUPRC) curve, the area under the ROC curve (AUCROC), precision, recall, F1 score, and balanced accuracy (BACC). All experiments were run in parallel on a 40-CPU research server with Ubuntu 16.04, and it took an average of 1.5 hours for embedding generation and 50 minutes for training on one configuration.

Dataset Split We employed two types of dataset splits for the model training and evaluation:

1. **Epitope to Antigen Surface Ratio:** This split ensures a similar distribution of epitope to non-epitope nodes across the training, validation, and test sets, with 1,383 complexes for training and 170 complexes each for validation and testing.
2. **Epitope Groups:** This split challenges the model to generalize to unseen epitope groups, with 641 unique antigens and 973 epitope groups. The dataset was divided into training, validation, and test sets with an 80/10/10 split, resulting in 1,383 complexes for training and 170 complexes each for validation and testing.

5 Results and Discussion

In this section, we report and discuss the results of our experiments on epitope node classification and bipartite link prediction tasks under epitope-to-antigen surface ratio and epitope group split settings using standard classification metrics.

5.1 Epitope Node Prediction

For epitope node classification, PLM embeddings demonstrated superior performance compared to classical encoding methods across both split strategies. Under epitope ratio dataset split (Table 1 - group a), AntiBERTy/ProtBERT embeddings configuration for an antibody-antigen pair achieved the highest MCC of 0.263 and AUCROC of 0.65, outperforming other embedding configurations. These results were consistent for both configurations of the decoder in our model, i.e., GCN and GCN with a fully connected linear layer. AntiBERTy consistently provided improved performance for antibody embeddings on the metrics of MCC, AUPRC, precision, and F1 score, while ProtBERT showed better performance for antigen embeddings across different combinations on these metrics. It is also observed that generating antigen embeddings using ESM-IF demonstrated very poor performance on all metrics except recall, while it provided average performance in the case of antibodies.

The more challenging epitope group-based split (Table 1 - group b) revealed interesting tradeoffs. The results were not consistent for a specific configuration of embeddings on the majority of evaluation metrics. In general, ESM2/ProtBERT led to the best results on the metrics of MCC, AUPRC, precision, F1 score, and balanced accuracy (BACC). Similar to the observations in the case of the epitope-ratio split, employing AntiBERTy for antibodies consistently led to better overall results on all metrics in the epitope group split. Consistent with the observations in [13], we also observed that the model struggled with the epitope group split setting compared to the epitope ratio because it couldn't generalize well on unseen epitopes and showed biases towards the training set. While our model with ESM2 embeddings attained the best results overall on these metrics, classical one-hot encoding surprisingly achieved the highest recall (0.556), suggesting simpler methods may capture broader structural patterns beneficial for unseen epitope detection. This dichotomy highlights the need for hybrid approaches combining PLM-derived semantics with geometric features.

Table 1: Performance of our model on the test set for epitope node prediction with epitope to antigen surface ratio (epitope ratio) and epitope group split strategies. The best values are highlighted in bold.

		Epitope Ratio Split							Epitope Group Split						
Algo.	Embedding (Ab/Ag)	MCC	AUPRC	AUCROC	Prec.	Recall	F1	BACC	MCC	AUPRC	AUCROC	Prec.	Recall	F1	BACC
GCN	AntiBERTy/ESM-IF	0.000	0.101	0.500	0.074	0.211	0.110	0.510	0.000	0.098	0.500	0.076	0.275	0.119	0.521
	AntiBERTy/ESM2	0.245	0.226	0.649	0.247	0.488	0.328	0.690	0.112	0.149	0.566	0.158	0.264	0.197	0.583
	AntiBERTy/ProtBERT	**0.263**	**0.251**	**0.650**	**0.281**	0.457	**0.348**	0.686	0.112	**0.153**	0.564	0.167	0.240	0.197	0.578
	ESM-IF/ESM-IF	0.000	0.101	0.500	0.068	1.000	0.127	0.500	0.000	0.098	0.500	0.065	1.000	0.122	0.500
	ESM-IF/ESM2	0.219	0.210	0.633	0.243	0.466	0.320	0.680	0.117	0.148	**0.573**	0.165	0.291	0.211	**0.594**
	ESM-IF/ProtBERT	0.232	0.225	0.640	0.253	0.459	0.326	0.680	0.107	0.152	0.560	0.162	0.231	0.190	0.574
	ESM2/ESM-IF	0.000	0.101	0.500	0.042	0.094	0.058	0.468	0.000	0.098	0.500	0.063	0.025	0.036	0.500
	ESM2/ESM2	0.228	0.211	0.643	0.231	0.504	0.317	0.691	0.112	0.151	0.566	0.157	0.252	0.193	0.579
	ESM2/ProtBERT	0.228	0.205	0.655	0.218	0.553	0.313	**0.705**	**0.117**	0.152	0.567	**0.181**	0.255	**0.212**	0.587
	ProtBERT/ESM-IF	0.000	0.101	0.500	0.068	**0.956**	0.127	0.501	0.000	0.098	0.500	0.061	0.445	0.108	0.485
	ProtBERT/ESM2	0.222	0.215	0.631	0.249	0.441	0.319	0.672	0.098	0.145	0.551	0.175	0.188	0.181	0.563
	ProtBERT/ProtBERT	0.236	0.225	0.642	0.259	0.461	0.332	0.682	0.094	0.144	0.552	0.172	0.195	0.183	0.565
	BLOSUM62/BLOSUM62	0.048	0.113	0.541	0.081	0.465	0.139	0.541	0.045	0.112	0.537	0.084	0.412	0.139	0.541
	One-Hot/One-Hot	0.049	0.113	0.543	0.079	0.533	0.138	0.542	0.044	0.108	0.542	0.076	**0.556**	0.134	0.543
GCN-L	AntiBERTy/ESM-IF	0.000	0.101	0.500	0.074	0.242	0.113	0.510	0.000	0.098	0.500	0.079	0.289	0.124	0.527
	AntiBERTy/ESM2	**0.256**	0.247	**0.650**	0.264	0.468	0.338	**0.687**	**0.123**	0.147	**0.569**	0.162	0.280	**0.205**	**0.590**
	AntiBERTy/ProtBERT	0.252	**0.251**	0.642	**0.283**	0.426	**0.340**	0.674	0.112	**0.153**	0.564	0.167	0.240	0.197	0.578
	ESM-IF/ESM-IF	0.000	0.101	0.500	0.068	**1.000**	0.127	0.500	0.000	0.098	0.500	0.065	1.000	0.122	0.500
	ESM-IF/ESM2	0.000	0.101	0.500	0.068	**1.000**	0.127	0.500	0.000	0.098	0.500	0.060	0.535	0.109	0.478
	ESM-IF/ProtBERT	0.219	0.218	0.634	0.240	0.472	0.318	0.682	0.099	0.150	0.557	0.155	0.239	0.188	0.574
	ESM2/ESM-IF	0.000	0.101	0.500	0.045	0.146	0.069	0.461	0.000	0.098	0.500	0.053	0.183	0.082	0.478
	ESM2/ESM2	0.220	0.203	0.641	0.222	**0.521**	0.311	0.694	0.060	0.071	0.430	0.032	0.024	0.074	0.038
	ESM2/ProtBERT	0.244	0.224	0.653	0.244	0.511	0.330	0.698	0.095	0.144	0.549	0.173	0.186	0.179	0.562
	ProtBERT/ESM-IF	0.000	0.101	0.500	0.056	0.293	0.094	0.466	0.000	0.098	0.500	0.060	0.535	0.109	0.478
	ProtBERT/ESM2	0.234	0.217	0.640	0.253	0.470	0.329	0.684	0.095	0.144	0.549	0.173	0.186	0.179	0.562
	ProtBERT/ProtBERT	0.216	0.204	0.634	0.238	0.480	0.318	0.684	0.115	0.151	0.564	**0.183**	0.234	0.205	0.581
	BLOSUM62/BLOSUM62	0.048	0.113	0.541	0.082	0.457	0.140	0.543	0.050	0.113	0.543	0.080	0.490	0.138	0.540
	One-Hot/One-Hot	0.045	0.112	0.539	0.078	0.509	0.136	0.537	0.046	0.109	0.544	0.076	0.542	0.134	0.543

5.2 Bipartite Link Prediction

The link prediction results (Table 2) demonstrate the task's inherent complexity, with maximum MCC scores of 0.126 and 0.054 for ratio and group splits, respectively. The performance gap between tasks suggests node classification benefits more from local residue features, while link prediction requires finer-grained interaction patterns

that current GCN architectures struggle to capture. Consistent with the epitope prediction task, the overall results for bipartite link prediction demonstrate that the AntiBERTy/ProtBERT embedding configuration achieves the best performance across almost all metrics for epitope-ratio split. Contrary to epitope node classification, we observed very high AUCROC scores of 0.852 for bipartite link prediction, where we were able to get a maximum AUCROC of 0.65. It was also observed that classical embedding methods such as BLOSUM62 and one-hot encoding showed degraded performance for both dataset splits and GCN decoder configurations on all metrics except AUCROC. Overall, these results suggest that GCN modules with a simple decoder performing the inner product of antibody and antigen embeddings provide consistent and best results compared to GCN with a linear layer as the decoder for both epitope node prediction and bipartite link prediction tasks. Consistent with the investigations in [10,13], our results demonstrate that the choice of the PLM embedding method impacts the model performance on the epitope prediction and bipartite link prediction tasks.

Table 2: Performance of our proposed model on bipartite link prediction task. The best values are highlighted in bold.

		Epitope Ratio Split							Epitope Group Split						
Algo.	Embedding (Ab/Ag)	MCC	AUPRC	AUCROC	Prec.	Recall	F1	BACC	MCC	AUPRC	AUCROC	Prec.	Recall	F1	BACC
GCN	AntiBERTy/ESM-IF	0.000	0.013	0.752	0.000	0.000	0.000	0.500	0.000	0.012	0.733	0.000	0.000	0.000	0.500
	AntiBERTy/ESM2	0.114	0.086	**0.852**	0.058	0.298	0.098	0.643	**0.054**	**0.038**	**0.796**	0.031	0.124	0.050	0.557
	AntiBERTy/ProtBERT	**0.126**	**0.100**	0.840	**0.063**	0.302	**0.104**	0.645	0.050	0.034	0.779	0.027	0.128	0.045	0.558
	ESM-IF/ESM-IF	0.000	0.005	0.496	0.000	0.000	0.000	0.500	0.000	0.004	0.510	0.000	0.000	0.000	0.500
	ESM-IF/ESM2	0.036	0.022	0.723	0.027	0.128	0.045	0.558	0.017	0.013	0.648	0.015	0.060	0.024	0.525
	ESM-IF/ProtBERT	0.042	0.023	0.713	0.028	0.164	0.047	0.574	0.015	0.012	0.644	0.012	0.071	0.021	0.528
	ESM2/ESM-IF	0.000	0.013	0.761	0.000	0.000	0.000	0.500	0.000	0.013	0.745	0.000	0.000	0.000	0.500
	ESM2/ESM2	0.100	0.064	0.843	0.048	0.308	0.083	0.646	0.051	0.032	0.779	0.027	**0.136**	0.045	**0.561**
	ESM2/ProtBERT	0.102	0.070	0.833	0.039	**0.373**	0.070	**0.674**	0.046	0.031	0.789	**0.034**	0.111	**0.052**	0.551
	ProtBERT/ESM-IF	0.000	0.013	0.763	0.000	0.000	0.000	0.500	0.000	0.014	0.750	0.000	0.000	0.000	0.500
	ProtBERT/ESM2	0.099	0.065	0.840	0.054	0.247	0.089	0.618	0.045	0.034	0.770	0.029	0.100	0.045	0.546
	ProtBERT/ProtBERT	0.113	0.078	0.827	0.053	0.312	0.090	0.649	0.043	0.028	0.768	0.030	0.108	0.047	0.549
	BLOSUM62/BLOSUM62	0.000	0.011	0.688	0.000	0.000	0.000	0.500	0.000	0.011	0.667	0.000	0.000	0.000	0.500
	One-Hot/One-Hot	0.000	0.011	0.687	0.000	0.000	0.000	0.500	0.000	0.011	0.668	0.000	0.000	0.000	0.500
GCN-L	AntiBERTy/ESM-IF	0.000	0.012	0.752	0.000	0.000	0.000	0.500	0.000	0.012	0.731	0.000	0.000	0.000	0.500
	AntiBERTy/ESM2	**0.121**	0.089	0.843	0.057	0.315	0.096	0.651	**0.057**	**0.037**	**0.807**	**0.037**	0.121	**0.056**	0.556
	AntiBERTy/ProtBERT	0.115	**0.090**	0.833	**0.059**	0.297	**0.098**	0.642	0.050	0.034	0.779	0.027	0.128	0.045	0.558
	ESM-IF/ESM-IF	0.000	0.004	0.504	0.000	0.000	0.000	0.500	0.000	0.004	0.490	0.000	0.000	0.000	0.500
	ESM-IF/ESM2	0.001	0.014	0.676	0.023	0.003	0.005	0.501	0.000	0.011	0.686	0.000	0.000	0.000	0.500
	ESM-IF/ProtBERT	0.045	0.023	0.708	0.027	0.178	0.046	0.580	0.020	0.013	0.641	0.014	0.083	0.024	0.534
	ESM2/ESM-IF	0.000	0.013	0.761	0.000	0.000	0.000	0.500	0.000	0.013	0.743	0.000	0.000	0.000	0.500
	ESM2/ESM2	0.091	0.054	0.845	0.045	0.278	0.077	0.631	0.045	0.028	0.767	0.029	0.106	0.046	0.548
	ESM2/ProtBERT	0.109	0.083	0.838	0.046	**0.344**	0.081	**0.662**	0.045	0.028	0.767	0.029	0.106	0.046	0.548
	ProtBERT/ESM-IF	0.000	0.013	0.765	0.000	0.000	0.000	0.500	0.000	0.014	0.749	0.000	0.000	0.000	0.500
	ProtBERT/ESM2	0.105	0.072	**0.850**	0.058	0.274	0.096	0.631	0.000	0.011	0.686	0.000	0.000	0.000	0.500
	ProtBERT/ProtBERT	0.098	0.067	0.835	0.049	0.288	0.083	0.637	0.053	0.033	0.771	0.031	**0.129**	0.050	**0.559**
	BLOSUM62/BLOSUM62	0.000	0.011	0.684	0.000	0.000	0.000	0.500	0.000	0.011	0.686	0.000	0.000	0.000	0.500
	One-Hot/One-Hot	0.000	0.011	0.680	0.000	0.000	0.000	0.500	0.000	0.012	0.669	0.000	0.000	0.000	0.500

5.3 Baselines Comparison

The comparative analysis of our proposed model against established baselines reveals significant performance improvements across multiple metrics (Table 3). Readers are referred to the related work (Section 2) for further description of these different baseline methods. For the epitope-to-antigen surface ratio split, our proposed model achieves an MCC of 0.263, substantially outperforming WALLE [13] and the next best model,

EpiPred [12]. Notably, WALLE's superior performance extends across all metrics, with particular strength in precision (0.281) and recall (0.457), indicating a balanced ability to identify true epitopes while minimizing false positives. The AUC-ROC of 0.650 further confirms our approach's robust epitope predictive capability, significantly surpassing ESMBind's 0.506. This improvement suggests that the integration of GCN with different PLMs captures critical structural and sequential information about antigens and antibodies that were previously not explored in detail.

Interestingly, the significant performance drop observed in the epitope groups split across all methods indicates a general challenge in generalizing to unseen epitopes. However, our model maintains its lead with an MCC of 0.077. While this represents a 70.7% decrease from the ratio split, it still outperforms other methods, many of which struggle to achieve positive MCC values in this setting. MaSIF-site [24] showed competitive performance in this split, particularly in precision, suggesting that its geometric deep learning approach captures some generalizable epitope features. This suggests that current approaches, including WALLE, may be overfitting to epitope patterns seen in training data, emphasizing the need for more diverse training sets and potentially more sophisticated regularization techniques.

Some graph-based epitope prediction approaches didn't enrich residue representations with PLM-based embeddings. For example, PECAN [20] originally employed a graph neural network (GNN) architecture with an attention layer to predict antibody-antigen binding sites on a dataset of 148 antibody-antigen complexes and achieved a recall of 0.67 on the epitope prediction task. Similarly, EPMP [11] used a GAT architecture with a fully connected antibody-antigen bipartite graph as input and achieved an AUCROC of 0.710 and AUPRC of 0.277. Our results demonstrate that specific PLM-based embeddings lead to improved performance on the epitope prediction task. Moreover, the performance gap between our proposed model and sequence-based methods like ESMFold and ESMBind underscores the importance of incorporating structural information for accurate epitope prediction [7,10]. ESMBind's near-random performance (AUC-ROC of 0.500 in the group split) highlights the limitations of purely sequence-based approaches for this task.

Table 3: Performance comparison of our proposed model with different baselines on two dataset split strategies. The values in bold represent the best values, while the dash represents that these metrics were not computed.

	Epitope Ratio Split					Epitope Group Split				
Algorithm	**MCC**	**Prec.**	**Recall**	**AUCROC**	**F1**	**MCC**	**Prec.**	**Recall**	**AUCROC**	**F1**
ESMBind	0.016	0.106	0.090	0.506	0.064	0.002	0.082	0.076	0.500	0.064
ESMFold	0.028	0.137	0.060	—	0.046	0.018	0.113	0.034	—	0.046
EpiPred	0.029	0.122	0.142	—	0.112	-0.006	0.089	0.158	—	0.112
MaSIF-site	0.037	0.125	0.114	—	0.128	0.046	**0.164**	0.174	—	0.128
WALLE	0.210	0.235	0.258	0.635	0.145	0.077	0.143	0.266	0.544	0.145
Our Approach	**0.263**	**0.281**	**0.457**	**0.650**	**0.348**	**0.123**	0.162	**0.280**	**0.569**	**0.205**

5.4 Limitations and Future Directions

Though our results demonstrate performance improvement for the epitope prediction task, there are limitations to our approach, and hence we note some potential future research directions. The authors in [10] developed a graph attention network (GAT) for antibody-agnostic epitope prediction and combined ESM2 and ESM-IF1 embeddings, demonstrating that the concatenation of these features improved the model performance. In the future, we will explore the integration of different PLM-based embedding methods, such as ESM2 with ProtBERT, for antibody-aware epitope prediction and bipartite link prediction tasks using separate GAT modules. Other deep learning models, such as bi-directional long short-term memory (BiLSTM [7]), have shown potential for predicting binding sites and need to be explored with concatenated PLM embeddings for the epitope prediction task. There is also immense potential in exploring other geometric deep learning models, such as diffusion-convolutional neural networks (DCNN [25]) for the epitope prediction task. We will also explore other protein representation approaches, such as point clouds and meshes [24], that have demonstrated success for protein binding site prediction [21].

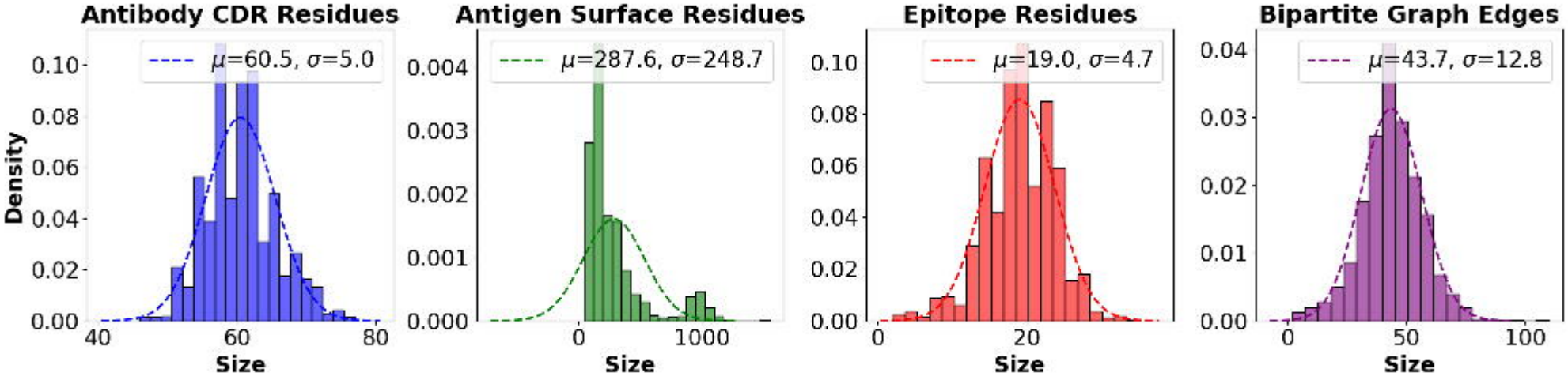
6 Conclusion

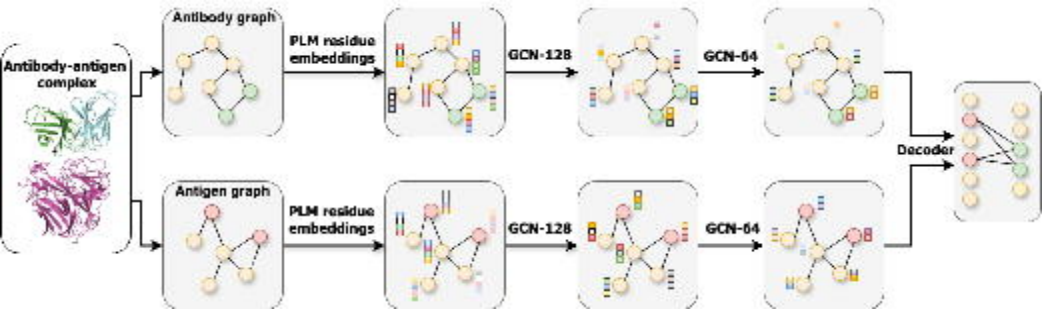
We extended a graph-based approach (WALLE) for antibody-aware epitope prediction on a novel benchmark dataset by exploring different PLM and classical embedding methods. We show that the use of PLM-based embeddings integrated with the graph-based models provides improved epitope prediction performance. Specifically, our experiments with different state-of-the-art PLM-based embedding methods show that using AntiBERTy for antibody embedding and ProtBERT for antigen embedding improves the epitope residue classification accuracy. We also discuss potential future research directions for improved epitope prediction.

References

1. R. A. Norman, F. Ambrosetti, A. M. Bonvin, L. J. Colwell, S. Kelm, S. Kumar, and K. Krawczyk, “Computational approaches to therapeutic antibody design: established methods and emerging trends,” *Briefings in bioinformatics*, vol. 21, no. 5, pp. 1549–1567, 2020.
2. S. Joubbi, A. Micheli, P. Milazzo, G. Maccari, G. Ciano, D. Cardamone, and D. Medini, “Antibody design using deep learning: from sequence and structure design to affinity maturation,” *Briefings in Bioinformatics*, vol. 25, no. 4, p. bbae307, 2024.
3. B. M. da Silva, D. B. Ascher, and D. E. Pires, “epitope1d: accurate taxonomy-aware b-cell linear epitope prediction,” *Briefings in Bioinformatics*, vol. 24, no. 3, p. bbad114, 2023.
4. S. R. Krishnan, D. Sharma, Y. Nazeer, M. Bose, T. Rajkumar, G. Jayaraman, N. Madaboosi, and M. M. Gromiha, “rabdesflow: a novel workflow for computational recombinant antibody design for healthcare engineering,” *Antibody Therapeutics*, vol. 7, no. 3, pp. 256–265, 2024.
5. A. M. Hummer, B. Abanades, and C. M. Deane, “Advances in computational structure-based antibody design,” *Current Opinion in Structural Biology*, vol. 74, p. 102379, 2022.
6. S. Fischman and Y. Ofran, “Computational design of antibodies,” *Current opinion in structural biology*, vol. 51, pp. 156–162, 2018.

7. Y. Zeng, Z. Wei, Q. Yuan, S. Chen, W. Yu, Y. Lu, J. Gao, and Y. Yang, "Identifying b-cell epitopes using alphafold2 predicted structures and pretrained language model," *Bioinformatics*, vol. 39, no. 4, p. btad187, 2023.
8. C. Wang, J. Wang, W. Song, G. Luo, and T. Jiang, "Episcan: accurate high-throughput mapping of antibody-specific epitopes using sequence information," *NPJ Systems Biology and Applications*, vol. 10, no. 1, p. 101, 2024.
9. Y. Fang, Y. Jiang, L. Wei, Q. Ma, Z. Ren, Q. Yuan, and D.-Q. Wei, "Deepprosite: structure-aware protein binding site prediction using esmfold and pretrained language model," *Bioinformatics*, vol. 39, no. 12, p. btad718, 2023.
10. S. Choi and D. Kim, "B cell epitope prediction by capturing spatial clustering property of the epitopes using graph attention network," *Scientific Reports*, vol. 14, no. 1, p. 27496, 2024.
11. A. Del Vecchio, A. Deac, P. Liò, and P. Veličković, "Neural message passing for joint paratope-epitope prediction," *arXiv preprint arXiv:2106.00757*, 2021.
12. K. Krawczyk, X. Liu, T. Baker, J. Shi, and C. M. Deane, "Improving b-cell epitope prediction and its application to global antibody-antigen docking," *Bioinformatics*, vol. 30, no. 16, pp. 2288–2294, 2014.
13. C. Liu, L. Denzler, Y. Chen, A. Martin, and B. Paige, "Asep: Benchmarking deep learning methods for antibody-specific epitope prediction," *arXiv preprint arXiv:2407.18184*, 2024.
14. Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, *et al.*, "Language models of protein sequences at the scale of evolution enable accurate structure prediction," *BioRxiv*, vol. 2022, p. 500902, 2022.
15. J. A. Ruffolo, J. J. Gray, and J. Sulam, "Deciphering antibody affinity maturation with language models and weakly supervised learning," *arXiv preprint arXiv:2112.07782*, 2021.
16. N. Brandes, D. Ofer, Y. Peleg, N. Rappoport, and M. Linial, "Proteinbert: a universal deep-learning model of protein sequence and function," *Bioinformatics*, vol. 38, no. 8, pp. 2102–2110, 2022.
17. C. Hsu, R. Verkuil, J. Liu, Z. Lin, B. Hie, T. Sercu, A. Lerer, and A. Rives, "Learning inverse folding from millions of predicted structures," in *International conference on machine learning*, pp. 8946–8970, PMLR, 2022.
18. A. Schreiber, "Esmbind and qbind: Lora, qlora, and esm-2 for predicting binding sites and post translational modification," *BioRxiv*, pp. 2023–11, 2023.
19. I. T. Desta, S. Kotelnikov, G. Jones, U. Ghani, M. Abyzov, Y. Kholodov, D. M. Standley, D. Beglov, S. Vajda, and D. Kozakov, "The cluspro abemap web server for the prediction of antibody epitopes," *Nature protocols*, vol. 18, no. 6, pp. 1814–1840, 2023.
20. S. Pittala and C. Bailey-Kellogg, "Learning context-aware structural representations to predict antigen and antibody binding interfaces," *Bioinformatics*, vol. 36, no. 13, pp. 3996–4003, 2020.
21. B. Dai and C. Bailey-Kellogg, "Protein interaction interface region prediction by geometric deep learning," *Bioinformatics*, vol. 37, pp. 2580–2588, 03 2021.
22. K. Jha, S. Saha, and H. Singh, "Prediction of protein–protein interaction using graph neural networks," *Scientific Reports*, vol. 12, no. 1, p. 8360, 2022.
23. S. Lu, Y. Li, Q. Ma, X. Nan, and S. Zhang, "A structure-based b-cell epitope prediction model through combing local and global features," *Frontiers in immunology*, vol. 13, p. 890943, 2022.
24. P. Gainza, F. Sverrisson, F. Monti, E. Rodola, D. Boscaini, M. M. Bronstein, and B. E. Correia, "Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning," *Nature Methods*, vol. 17, no. 2, pp. 184–192, 2020.
25. J. Atwood and D. Towsley, "Diffusion-convolutional neural networks," *Advances in neural information processing systems*, vol. 29, 2016.


Antibody CDR Residues
μ=60.5, σ=5.0
Density
Size
Antigen Surface Residues
μ=287.6, σ=248.7
Size
Epitope Residues
μ=19.0, σ=4.7
Size
Bipartite Graph Edges
μ=43.7, σ=12.8
Size


Antibody-antigen complex
Antibody graph
PLM residue embeddings
GCN-128
GCN-64
Antigen graph
PLM residue embeddings
GCN-128
GCN-64
Decoder