bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452188; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A Structure-based B-cell Epitope Prediction Model Through Combing Local and Global Features

Shuai Lu,[1,2,3] Yuguang Li,[1,2,3] Xiaofei Nan,[1,2,3] and Shoutao Zhang[1,2,3]

[1] *School of Information Engineering, Zhengzhou University, Zhengzhou, Henan, 450001, China.*
[2] *College of Economics and Management, Zhengzhou University of Light Industry, Zhengzhou, Henan, 450001, China.*
[3] *School of Life Science, Zhengzhou University, Zhengzhou, Henan, 450001, China.*

B-cell epitopes (BCEs) are a set of specific sites on the surface of an antigen that binds to an antibody produced by B-cell. The recognition of epitopes is a major challenge for drug design and vaccines development. Compared with experimental methods, computational approaches have strong potential for epitope prediction at much lower cost. Moreover, most of the currently methods focus on using local information around target amino acid residue for BCEs prediction without taking the global information of the whole antigen sequence into consideration.We propose a novel deep leaning method thorough combing local features and global features for BCEs prediction. In our model, two parallel modules are built to extract local and global features from the antigen separately. For local features, we use graph convolutional networks to capture information of spatial neighbors of a target amino acid residue. For global features, Attention based Bidirectional Long Short-Term Memory networks (Att-BLTM) are applied to extract information from the whole antigen sequence. Then the local and global features are combined to predict BCEs. The experiments show that the proposed method achieves superior performance over the state-of-the-art BCEs prediction methods on benchmark datasets. Also, we compare the performance differences between data with or without global features. The experimental results show that global features play an important role in BCEs prediction.

I. INTRODUCTION

The humoral immune system protects the body from foreign objects like bacteria and viruses by developing B-cells and producing antibodies[1]. Antibodies play a crucial role in immune response through recognizing and binding the disease-causing agents, called antigen. B-cell epitopes (BCEs) are a certain region on the antigen surface that are bound by an antibody[2]. BCEs of protein antigens can be roughly classified into two categories, linear and conformational[3]. Linear BCEs consist of residues that are contiguous in the antigen primary sequence, while the conformational BCEs comprise residues which are not contiguous in sequence but folding together in three-dimensional structure space. About 10% of BCEs are linear and about 90% are conformational[4]. In our study, we focus on conformational BCEs of protein antigens.

The localization and identification of epitopes is of great importance for the development of vaccines and for the design of therapeutic antibodies[5], [6]. However, traditional experimental methods to identify BCEs are still expensive and time-consuming[7]. Therefore, great efforts for computational approaches based on machine learning algorithms have been developed to predict BCEs. These approaches can be divided in two categories: sequence-based and structure-based methods. As the name implies, the sequence-based approaches predict BCEs only based on the antigen sequence, while the structure-based approaches also consider its structural features. Currently, various structure-based predictors have been developed to predict and analyze BCEs including BeTop[8], Bpredictor[9], DiscoTope 2.0[10], CE-KEG[11], CeePre[12], EpiPred[13], ASE_Pred[14] and PECAN[15].

Some of those methods improve model performance by introducing novel features such as statistical features in BeTop, thick surface patch in Bpredictor, new spatial neighborhood definition and half-sphere exposure in DiscoTope2.0, knowledge-based energy and geometrical neighboring residue contents in CE-KEG, B factor in CeePre and surface patches in ASE_Pred. Except novel features, antibody structure and powerful model also improve the results of BCEs prediction. Epipred utilizes antibody structure information to annotate the epitope region and improves global docking results. PECAN represents antigen or antibody structure as a graph and employ graph convolution operation on them to make aggregation of spatial neighboring residues. An additional attention layer is used to encode the context of the partner antibody in PECAN. However, all the currently structure-based BCEs prediction methods only use local information around target amino acid residue without considering the global information of the whole antigen sequence.

Global features have been proved to be effective in some biology sequence analysis models such as protein-protein interaction sites prediction model DeepPPISP[16] and protein phosphorylation sites prediction model DeepPSP[17]. However, which model is used determines the effect of global features. DeepPPISP utilizes TextCNNs processing the whole protein sequence for protein-protein interaction sites prediction. DeepPSP employs SENet blocks and Bi-LSTM blocks to extract the global features for protein phosphorylation sites. In our study, we take advantage of the Attention based Bidirectional Long Short-Term Memory(Att-BLTM) networks which are first introduced for relation classification

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452188; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in the field of natural language processing (NLP)[18]. Att-BLSTM networks are also employed for some chemical and biomedical text processing tasks including chemical named entity recognition[19] and biomedical event extraction[20]. Given the excellent performance of Att-BLSTM, we combine it with the novel deep learning model Graph Convolution Networks(GCNs)[21] for BCEs prediction.

In this study, we propose a structure-based BCEs prediction model utilizing both antigen local features and global features. By combining Att-BLSTM and GCNs, both local and global features are used in our model to improve its prediction performance. We implement our model on some public datasets and the results show that global features can provide useful information for BCEs prediction.

II. MATERIALS AND METHODS

A. Datasets

We use the same antibody-antigen complexes as PECAN[15]. Those complexes are from two separate datasets, one from EpiPred[13], and the other from Docking Benchmarking Dataset (DBD) v5[22]. There are 148 antibody–antigen complexes from EpiPred[13], 118 for training and 30 for testing and all of them shared no more than 90% pairwise sequence identity. In order to construct a separate validation set, [15] filters the antibody–antigen complexes in the Docking Benchmarking Dataset (DBD) v5[22] to guarantee that all antigens in the validation set had no more than 25% pairwise sequence identity to those in the testing set. Together 162 complexes are selected. 103 complexes are used for training, 29 for validation and 30 for testing. Same as works in[13], [15], residues are labeled as part of the epitope if they have any heavy atom within 4.5 Å away from any heavy atom on the antibody. The size of datasets and number of epitopes are shown in Table 1.

TABLE I: Sumarry of datasets

Datasets	NO. of Complexes	NO. of Epitopes	NO. of non-Epitopes
TrainingSet	103	2708	19567
Validation	29	839	5553
TestingSet	30	758	6434

B. Input features representation

For global features, we construct the input antigen sequence as a set of sequential amino acid residues:

$$S = [r_1, r_2, r_3, \cdots, r_i, \cdots, r_l]^T, S \in R^{(l*d)} \quad (1)$$

where each residue is represented as a vector $r_i \in R^d$ corresponding to the i-th residue in the antigen sequence, l is the antigen sequence length, and d is the residue feature dimension. For local features, each antigen structure is represented as a graph as related studies in [15] and [23]. The amino acid residue is a node in the protein graph whose features represent its properties. For residue r_i, the local environment consists of k spatial neighboring residues, and constructs its receptive field N_i :

$$N_i = \{n_1, \cdots, n_k\} \quad (2)$$

$\{r_{n_1}, \cdots, r_{n_k}\}$ are the neighbors of residue r_i which define the operation field of the graph convolution.

In this study, the residue node feature is a 128-dimension vector encoding important properties used as in our earlier work [24]. All those residue node features can be divided into two classes: sequence-based and structure-based. Sequence-based features consist of the one-hot encoding of the amino acid residue type, seven physicochemical parameters[25] and a conservation profile returned by running PSI-BLAST[26] against nr database[27]. The structure-based features are calculated for each antigen structure isolated from the antibody-antigen complex by DSSP[28], MSMS[29], PSAIA[30] and Biopython[31]. The residue edge features between two residues r_i and r_j are representing as e_ij. e_ij reflects the spatial relationships including the distance and angle between residue pair r_i and r_j and it is computed by their C_a[23].

C. Model architecture

Our model solves a binary classification problem: judging an antigen residue binding to its partner antibody or not. As shown in Fig. 1, our model consists of two parallel parts: GCNs and Att-BLSTM networks. The former captures local features of target antigen residue from its spatial neighbors, and the latter extracts global features from the whole antigen sequence. The outputs are concatenated and fed to two linked fully connected networks to predict the binding probability for each antigen residue.

1. *Graph convolutional networks*

Each protein is represented as a graph, and a residue is a node in the graph. We perform graph convolution operation on the receptive field of each residue. In our study, we utilize two graph convolution operators as previous work[23]:

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452188; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

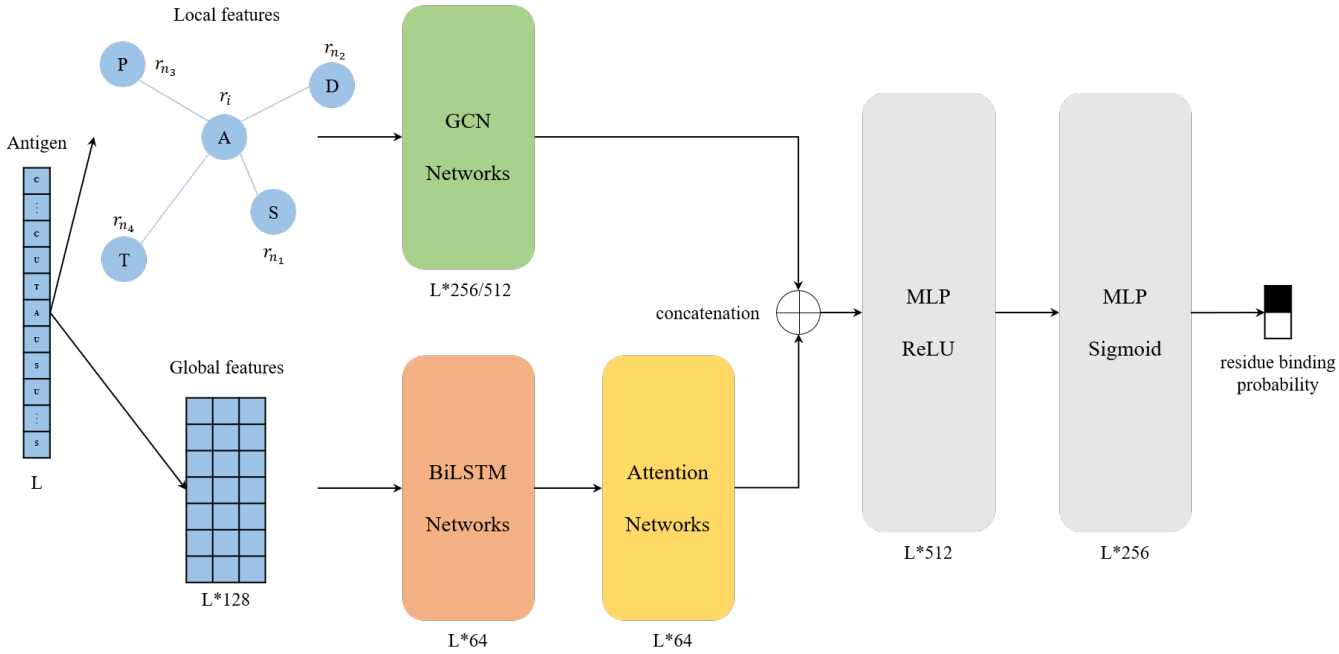

FIG. 1: Model architecture in this study.

$$z_i = \sigma(W_t + \frac{1}{|N_i|} + \sum_{j \in N_i} W_n r_j + b_n) \quad (3)$$

$$z_i = \sigma(W_t + \frac{1}{|N_i|} + \sum_{j \in N_i} W_n r_j + \sum_{j \in N_i} W_e e_{ij} + b_{ne}) \quad (4)$$

where N_i is the receptive field, i.e. a set of neighbors of target residue r_i, W_t is the weight matrix associated with the target node, W_n is the weight matrix associated with neighboring nodes, σ is a non-linear activation function, and b_n is a bias vector. Formula 3 groups the node information in receptive filed. Formula (4) utilizes not only node features but also edge feature between two residues, where W_e is the weight matrix associated with edge feature, e_ij represents the edge feature between residue r_i and r_j, and b_ne is a vector of biases.

2. *Attention-based bidirectional Long Short-Term Memory networks*

Besides local features, global features are crucial in BCEs prediction as well. In our work, Attention-based Bidirectional Long Short-Term Memory(Att-BLSTM) networks are used to capture global sequence information of input antigen sequence. Att-BLSTM is first proposed for relation classification which is a natural language processing (NLP) task[18]. It captures the most important semantic information in a sentence and achieves the state-of-the-art performance on a public dataset. Currently, Att-BLSTM has been used for processing chemical and biomedical text[19], [20]. However, its advantage has not been exploited in biology sequence analysis such as BCEs prediction.

Fig. 2 shows the architecture of Att-BLSTM. The input antigen sequence is represented as a set of residues $S = [r_1, r_2, r_3, \cdots, r_i, \cdots, r_l]^T, S \in R^{(l*d)}$. S is then fed into a Long Short-Term Memory (LSTM) network which learns long-range dependencies in a sequence[32]. Typically, the structure of an LSTM unit at each time t is calculated by the following formulas:

$$i_t = \sigma(W_i * [h_{t-1}, r_t] + b_i) \quad (5)$$

$$f_t = \sigma(W_f * [h_{t-1}, r_t] + b_f) \quad (6)$$

$$o_t = \sigma(W_o * [h_{t-1}, r_t] + b_o) \quad (7)$$

$$c_t = f_t * c_{t-1} + i_t * (tanh(W_c * [h_{t-1}, r_t] + b_c)) \quad (8)$$

$$h_t = o_t * tanh(c_t) \quad (9)$$

where tanh is the element-wise hyperbolic tangent, σ is the logistic sigmoid function, r_t, h_{t-1} and c_{t-1} are inputs, and h_t and c_t are outputs. There are three gates

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452188; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

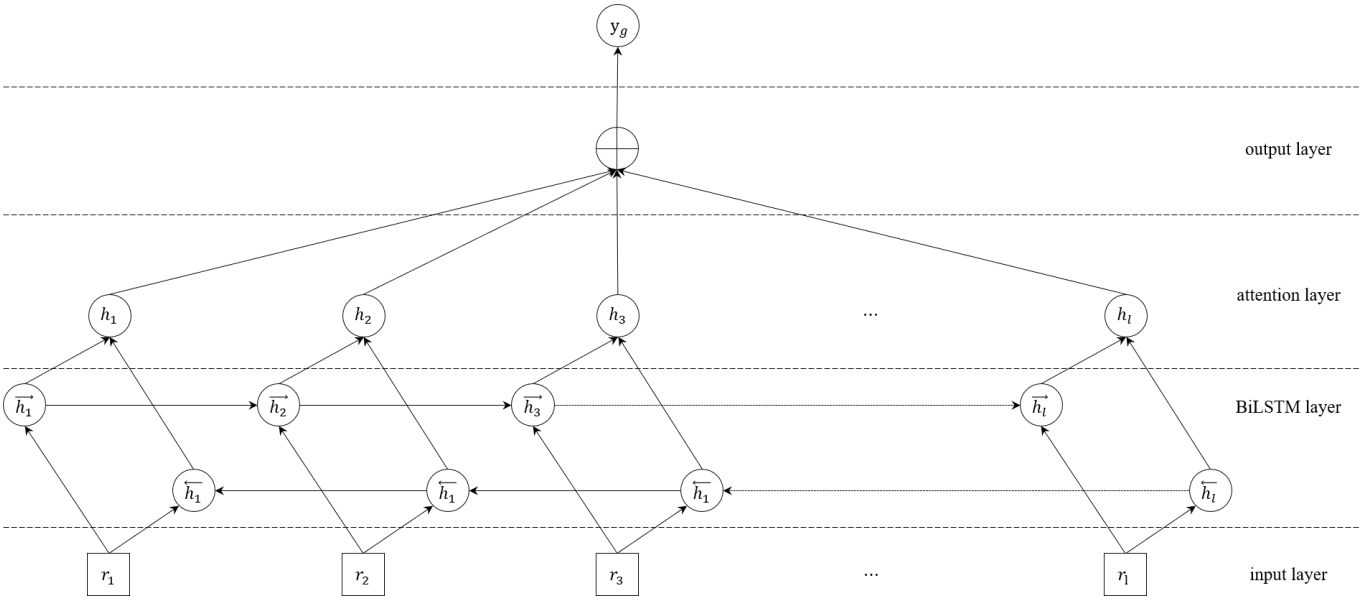

FIG. 2: Model architecture of Att-BLSTM.

consisting of one input gate i_t with corresponding weight matrix W_i, and a bias b_i; one forget gate f_t with corresponding weight matrix W_f, and a bias b_f; one output gate o_t with corresponding weight matrix W_o, and a bias b_o.

Bidirectional LSTM(BiLSTM) can learn forward and backward information of input sequence. As shown in Figure 2, the networks contain two sub-networks for the left and right sequence contexts. For the i-th residue in the input antigen sequence, we combine the forward pass output $\overrightarrow{h_i}$ and backward pass output $\overleftarrow{h_i}$ by concatenating them:

$$h'_i = [\overrightarrow{h_i} \oplus \overleftarrow{h_i}] \tag{10}$$

The output of BiLSTM layer is matrix H which consists of all output vectors of input antigen residues: $H = [h'_1, h'_2, h'_3, \ldots, h'_i, \ldots, h'_l]^T, H \in R^{(l*2d)}$, where l is the input antigen sequence length, and d is the residue features dimension.

The attention layer of Att-BLSTM networks employs attention mechanism which has been used in a lot of biology tasks ranging from compound–protein interaction prediction[33], paratope prediction[15] and protein structure prediction[34]. In attention layer of Att-BLSTM, the novel representation S' as well as the output y_g of the input antigen sequence is formed by a weighted sum of those output vectors H:

$$M = tanh(H) \tag{11}$$

$$\alpha = softmax(W_\mu M) \tag{12}$$

$$y_g = S' = H\alpha^T \tag{13}$$

3. *Performance evaluation*

As shown in Fig. 1 and 2, the local features z_i extracted by GCNs and the global features y_g derived from Att-BLSTM networks are concatenated. And then, they are fed to two linked fully-connected networks. The calculation of probability y_i for each input antigen residue belonging to epitope is shown as:

$$y_i = f(W(y_g \oplus z_i) + b) \tag{14}$$

D. Performance evaluation

Most structure-based methods use precision and recall as evaluation metrics. Our model predicts a probability for each antigen residue, and precision and recall are calculated by taking probability threshold of 0.5. Structure-based BCEs predictors DiscoTope 2.0[10] and EpiPred[13] report precision and recall of each antigen in testing set. And PECAN[15] calculates averaged precision, recall and area under the precision recall curve (AUC PR) over all antigens in the testing set. To compare with those state-of-the-art structure-based BCEs methods, we use evaluation metrics including averaged precision, recall and AUC PR over all antigens in testing set, as well as the precision and recall values of each testing antigen.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452188; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

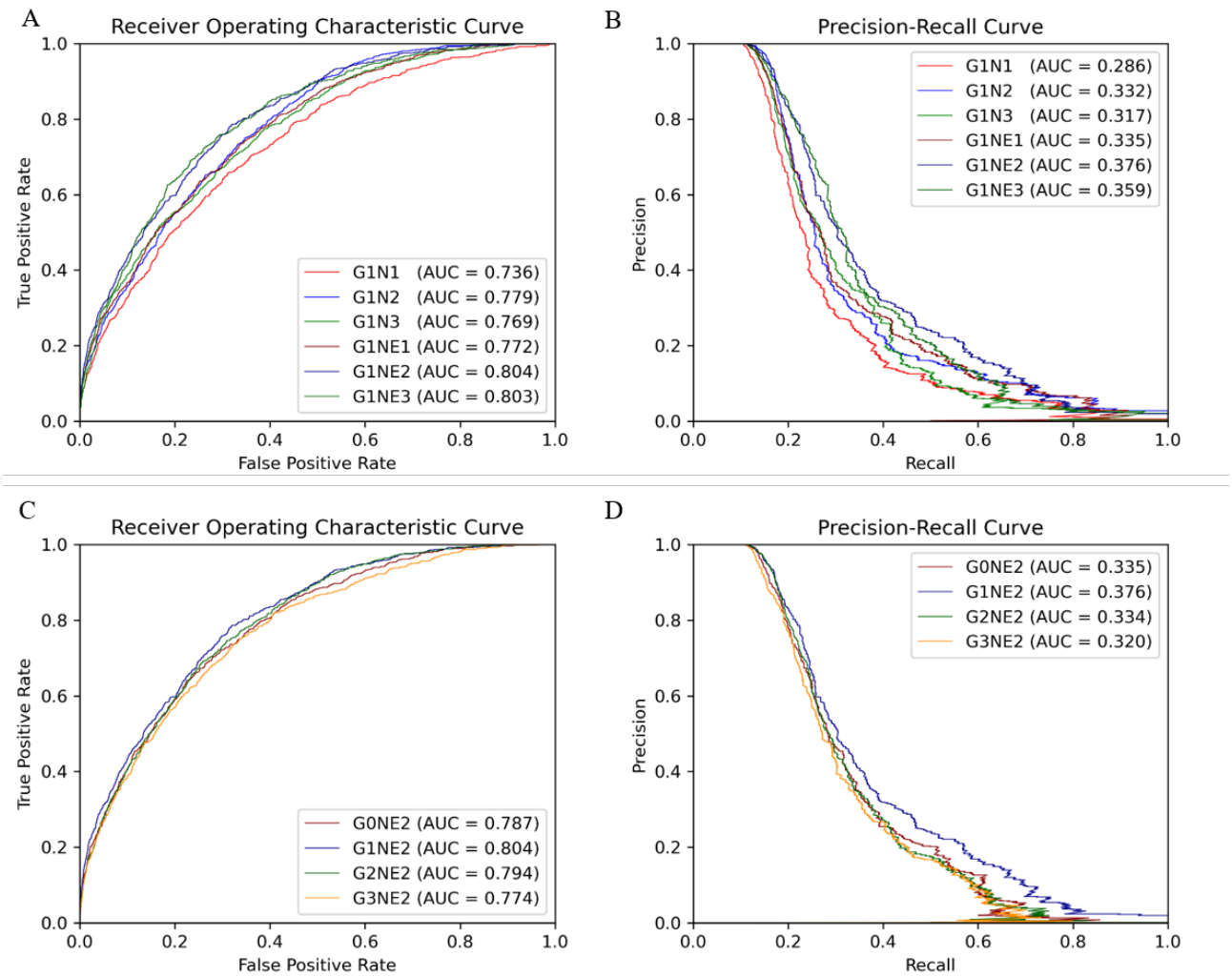

FIG. 3: ROC and PR curve of different network combinations

In order to evaluate the effects of local and global features in our method, we take advantage of the area under the receiver operating characteristics curve (AUC ROC) and AUC PR. AUC PR is more sensitive than AUC ROC on imbalanced data. And, the datasets used for BCEs prediction are roughly 90% negative class. Therefore, we take AUC PR as the most import metric for model evaluation[35].

E. Implementation details

We implement our model using PyTorch v1.4. The training details of these neural networks are as follows: optimization: Momentum optimizer with Nesterov accelerated gradients; learning rate: 0.001; batch size: 32; dropout: 0.5; spatial neighbors in the graph: 20; number of LSTM layers in Att-BLSTM networks: 1, 2 or 3; number of graph convolution networks layers: 1, 2 or 3. Training time of each epoch varies from roughly 1 to3 minutes depending on network depth, using a single NVIDIA RTX2080 GPU.

For each combination, networks are trained until the performance on the validation set stops improving or for a maximum of 250 epochs. Graph convolution networks have the following number of filters for 1, 2 and 3 layers, respectively: (256), (256, 512), (256, 256, 512). All weight matrices are initialized as and biases are set to zero[23].

III. RESULTS

A. The effects of different network combinations

In this section, we focus on which network combinations are most effective. The AUC ROC and AUC PR are shown in Figure 3, where label G denotes the model processing global features and the number after G indicates the model depth, for instance, G1 represents 1-layer Att-

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452188; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

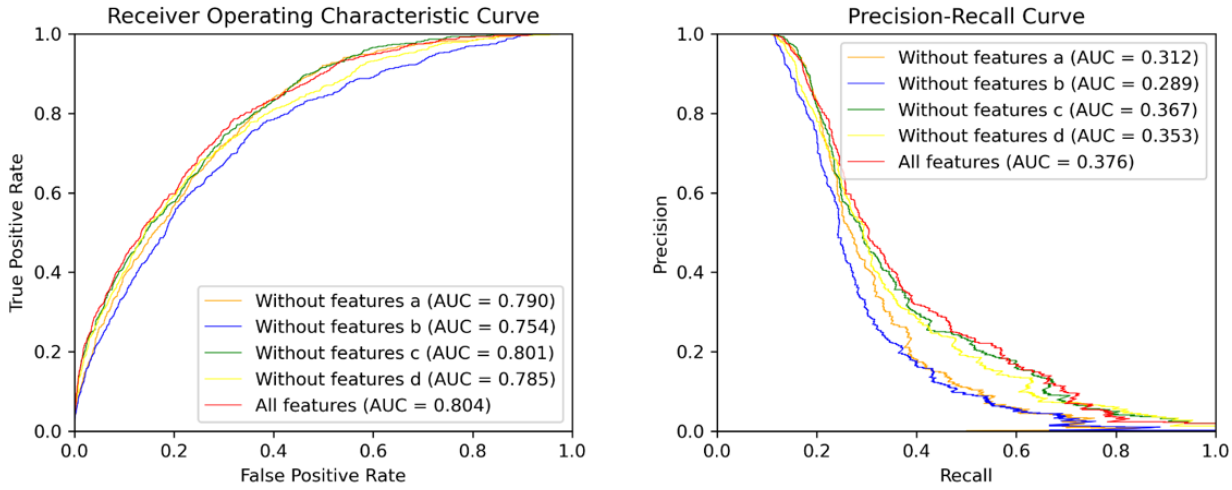

FIG. 4: ROC and PR curve of different combinations of input features

BLSTM. And N denotes residue Node features and NE denotes both residue Node and Edge features are used in GCNs. The number behind N or NE indicates the depth of GCNs. For example, G1NE2 represents model with 1-layer Att-BLSTM and 2-layer GCNs using residue node and edge features.

First, we train our model of 1-layer Att-BLSTM with varying GCNs depths with or without residue edge features. From figure 3A and 3B, we observe that the 2-layer GCNs with residue edge features perform best (AUC ROC = 0.804, AUC PR = 0.376). This draws the same conclusion with our earlier work for antibody paratope prediction[24]. We also find that residue edge features can always provide better performance as the GCNs depths vary. The same results are found in protein interface prediction task using GCNs as well[23].

Second, 2-layer GCNs and Att-BLSTM networks of different depths are combined in our model. Figure 3C and 3D show the performance evaluated by AUC ROC and AUC PR. It can be found that the combination of 1-layer Att-BLSTM network and 2-layer GCNs with residue edge features still has the best results. In general, the deeper the Att-BLSTM networks grow, the results get worse. As discussed in DeepPPISP[16], global features may cover the relationships among residues of longer distances. However, those relationships might become weaken when the Att-BLSTM networks grow deeper. In summary, our model with 2-layes GCNs and 1-layer Att-BLSTM network performs best, and it is the proposed model in this paper and used for comparison with competing methods in the following sections.

B. The effects of global features

The global feature has been shown to improve the performance of protein-protein interaction sites prediction in DeepPPISP[16] and protein phosphorylation sites prediction in DeepPSP[17]. In order to verify whether global features are effective in BCEs prediction as well, we remove the Att-BLSTM networks in our model for comparison. As shown in Figure 3C and 3D, label G0 means there is only GCNs in our model, and no global features are used. Without global features, the AUC ROC is 0.787, which is lower than the proposed model G1NE2 (also lower thanG2NE2, but slightly better than G3NE2). Without global features, the AUC PR is 0.335, which is significantly worse than the proposed model (but slightly better than G2NE2 and G3NE2). The model without global features performs worse on both AUC ROC and AUC PR metrics than our proposed model. Therefore, global features improve the performance of our model for BCEs prediction.

However, models with global features not always outperform model without global features in our experiment. Similar observation has been found in DeepPSP[17], but DeepPPISP[16] reaches a contrary conclusion. This situation might be caused by different models processing global features. In DeepPPISP, a simple fully-connected network is used, and in DeepPSP, SENet blocks and Bi-LSTM blocks are used.

C. The effects of different types of input features

Different types of input features (sequence and structure-based) play different roles in our model. The input features consist of four types: residue type one-hot encoding (a), profile features (b), seven physicochemical parameters (c) and structural features (d). To discover what role each feature type plays in our method, we delete each input feature type and compare their performances on our proposed model (G1NE2, i.e. 1-layer Att-BLSTM network and 2-layersGCNs with residue edge

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452188; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

features). Figure 4 shows the experimental results. As Figure 4A shows, the AUC ROC without features b is 0.754, significantly lower than the best performance 0.804. The AUC PR without features b drops biggest from 0.376 (all features) to 0.289. It indicates that profile features (feature type b) are most important in our model for BCEs prediction. The model using all the features still performs best on both AUC ROC and AUC PR metrics.

D. Comparison with competing methods

TABLE II: Performances of BCEs prediction methods

Method	AUC PR	Precision	Recall
EpiPred	-	0.136	0.436
DiscoTope	-	0.214	0.110
PECAN	0.212 ± 0.007	0.154	**0.691**
Our Method	**0.457 ± 0.001**	**0.638**	0.546

To evaluate the performance of our method for BCEs prediction, we compare our proposed model with three competing structure-based BCEs prediction methods: DiscoTope[10], EpiPred[13] and PECAN[15]. Note that these methods all used local features but did not consider global features. Table 2 shows the experimental results of averaged AUC PR, precision and recall on testing set of our proposed model. The results on three competing methods are taken from [15]. Although our model gets lower recall than PECAN, it is higher than all other competing methods on precision and makes distinct improvement on AUC PR.

TABLE III: The summarizing results on testing set

PDB	Ag size	EpiPred		DiscoTope 2.0		Our Method	
		Precision	Recall	Precision	Recall	Precision	Recall
4hj0	92	0.32	**0.9**	0	0	**0.523**	0.513
1tzh	94	0.01	0.06	0.73	**0.87**	**0.579**	0.516
4am0	96	0.13	**0.7**	0.33	0.2	**0.936**	0.571
2ih3	97	0.16	**0.64**	0	0	**0.919**	0.605
4i77	97	0.23	0.55	0	0	**0.871**	**0.629**
3q1s	113	0.19	**0.81**	0	0	**0.662**	0.632
1p2c	129	0	0	**1**	0.05	0.715	**0.607**
4ht1	131	0.05	0.14	0	0	**0.374**	**0.500**
3ab0	136	0.33	**0.73**	0	0	**0.401**	0.500
1v7m	145	0.26	**0.77**	0	0	**0.825**	0.576
4g3y	148	0.03	0.08	**1**	0.17	0.730	**0.617**
2vxt	156	0.04	0.09	0.47	0.36	**0.717**	**0.546**
3u9p	169	0.31	**1**	0.06	0.05	**0.430**	0.486
3o2d	178	0.32	**0.64**	0	0	**0.877**	0.635
1fns	196	0	0	**1**	0.07	0.457	**0.500**
3ma9	197	0	0	0	0	**0.522**	**0.505**
3rvv	223	0.25	**0.93**	0.15	0.17	**0.448**	0.498
3raj	230	0	0	0	0	**0.453**	**0.488**
1nfd	239	0.07	0.23	**0.92**	0.7	0.626	0.538
3i50	273	0	0	0	0	**0.800**	**0.548**
3gjf	276	0.15	**0.66**	0.05	0.11	**0.589**	0.537
3liz	329	0.26	**0.68**	0	0	**0.454**	0.498
3pgf	358	0.02	0.04	0	0	**0.566**	**0.507**
3zkm	375	0.32	**0.88**	0	0	**0.954**	0.528
3r1g	381	0.37	**1**	0	0	**0.712**	0.515
4jr9	409	0.19	**0.85**	0.46	0.5	**0.709**	0.709
4ene	442	0	0	0	0	**0.481**	**0.500**
3o0r	449	0.08	**0.7**	0	0	**0.614**	0.576
3t3p	453	0	0	0.25	0.04	**0.720**	**0.517**
1n8z	581	0	0	0	0	**0.481**	**0.487**

We also compare the results of each antigen in testing set with DiscoTope 2.0 and EpiPred. The results presented of DiscoTope 2.0 and EpiPred in Table 3 are taken form [13]. The values in bold indicate the best prediction result. We find that our model achieves best precision on 26 antigens and best recall on 13 antigens of all 30 antigens in testing set. We also observe that our model produces usable prediction even for the long antigen target.

IV. CONCLUSION

Accurate prediction of BCEs is helpful for understanding the basis of immune interaction and is beneficial to therapeutic design. In this work, we propose a novel deep learning framework combining local and global features which are extracted from antigen sequence and structure to predict BCEs. GCNs are used for capturing the local features of a target residue. Att-BLSTM networks are used to extract global features, which figure the relationship between a target residue and the whole antigen. We employ our model on a public and popular dataset and the results show improvement of BCEs prediction. Moreover, our results declare that the global features are useful for improving the prediction of BCEs.

Though our method outperforms other competing computational methods for BCEs prediction, it also has some disadvantages. The first one is that our predictor needs antigen structure as it takes structure-based residue features as input. The second one is that it consumes long computer time because PSI-BLAST[26] needs to be performed at the stage of extracting residue features.

In this study, we show that combing local and global features can be useful for BCEs prediction. In the future, we would further improve BCEs prediction by expanding

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452188; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the training set.

[1] E. D. Getzoff, J. A. Tainer, R. A. Lerner, and H. M. Geysen, "The Chemistry and Mechanism of Antibody Binding to Protein Antigens," Adv. Immunol., vol. 43, no. C, pp. 1–98, Jan. 1988, doi: 10.1016/S0065-2776(08)60363-6.

[2] S. W. Michnick and S. S. Sidhu, "Submitting antibodies to binding arbitration," Nature Chemical Biology, vol. 4, no. 6. Nature Publishing Group, pp. 326–329, 2008, doi: 10.1038/nchembio0608-326.

[3] D. J. Barlow, M. S. Edwards, and J. M. Thornton, "Continuous and discontinuous protein antigenic determinants," Nature, vol. 322, no. 6081, pp. 747–748, 1986, doi: 10.1038/322747a0.

[4] S. E. C. Caoili, "Hybrid methods for B-cell epitope prediction approaches to the development and utilization of computational tools for practical applications," Methods Mol. Biol., vol. 1184, pp. 245–283, 2014, doi: 10.1007/978-1-4939-1115-8_14.

[5] J. M. Gershoni, A. Roitburd-Berman, D. D. Siman-Tov, N. T. Freund, and Y. Weiss, "Epitope mapping: The first step in developing epitope-based vaccines," BioDrugs, vol. 21, no. 3. BioDrugs, pp. 145–156, 2007, doi: 10.2165/00063030-200721030-00002.

[6] A. C. Chan and P. J. Carter, "Therapeutic antibodies for autoimmunity and inflammation," Nature Reviews Immunology, vol. 10, no. 5. Nature Publishing Group, pp. 301–316, May 2010, doi: 10.1038/nri2761.

[7] W. M. Abbott, M. M. Damschroder, and D. C. Lowe, "Current approaches to fine mapping of antigen-antibody interactions," Immunology, vol. 142, no. 4, pp. 526–535, 2014, doi: 10.1111/imm.12284.

[8] L. Zhao, L. Wong, L. Lu, S. C. H. Hoi, and J. Li, "B-cell epitope prediction through a graph model," BMC Bioinformatics, vol. 13, no. S20, pp. 1–12, 2012, doi: 10.1186/1471-2105-13-S17-S20.

[9] W. Zhang, Y. Xiong, M. Zhao, H. Zou, X. Ye, and J. Liu, "Prediction of conformational B-cell epitopes from 3D structures by random forests with a distance-based feature," BMC Bioinformatics, vol. 12, 2011, doi: 10.1186/1471-2105-12-341.

[10] J. V. Kringelum, C. Lundegaard, O. Lund, and M. Nielsen, "Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking," PLoS Comput. Biol., vol. 8, no. 12, Dec. 2012, doi: 10.1371/journal.pcbi.1002829.

[11] Y. T. Lo, T. W. Pai, W. K. Wu, and H. T. Chang, "Prediction of conformational epitopes with the use of a knowledge-based energy function and geometrically related neighboring residue characteristics," BMC Bioinformatics, vol. 14, no. SUPPL4, 2013, doi: 10.1186/1471-2105-14-S4-S3.

[12] J. Ren, Q. Liu, J. Ellis, and J. Li, "Tertiary structure-based prediction of conformational B-cell epitopes through B factors," Bioinformatics, vol. 30, no. 12, pp. 264–273, 2014, doi: 10.1093/bioinformatics/btu281.

[13] 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, doi: 10.1093/bioinformatics/btu190.

[14] M. C. Jespersen, S. Mahajan, B. Peters, and M. Nielsen, "Antibody Specific B-Cell Epitope Predictions: Leveraging Information From Antibody-Antigen Protein Complexes," Front. Immunlogy, vol. 10, no. 298, pp. 1–10, 2019, doi: 10.3389/fimmu.2019.00298.

[15] 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, doi: 10.1093/bioinformatics/btaa263.

[16] M. Zeng, F. Zhang, F. X. Wu, Y. Li, J. Wang, and M. Li, "Protein-protein interaction site prediction through combining local and global features with deep neural networks," Bioinformatics, vol. 36, no. 4, pp. 1114–1120, 2020, doi: 10.1093/bioinformatics/btz699.

[17] L. Guo et al., "DeepPSP: A Global-Local Information-Based Deep Neural Network for the Prediction of Protein Phosphorylation Sites," J. Proteome Res., vol. 20, no. 1, pp. 346–356, 2021, doi: 10.1021/acs.jproteome.0c00431.

[18] P. Zhou et al., "Attention-based bidirectional long short-term memory networks for relation classification," in Annual Meeting of the Association for Computational Linguistics, 2016, pp. 207–212, doi: 10.18653/v1/p16-2034.

[19] L. Luo et al., "An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition," Bioinformatics, vol. 34, no. 8, pp. 1381–1388, 2018, doi: 10.1093/bioinformatics/btx761.

[20] L. Li, J. Wan, J. Zheng, and J. Wang, "Biomedical event extraction based on GRU integrating attention mechanism," BMC Bioinformatics, vol. 19, no. Suppl 9, 2018, doi: 10.1186/s12859-018-2275-2.

[21] Kipf, "Semi-supervised classfication with graph convolutional networks," in International Conference on Learning Representations, 2017, pp. 1–10.

[22] T. Vreven et al., "Updates to the Integrated Protein-Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2," J. Mol. Biol., vol. 427, no. 19, pp. 3031–3041, Sep. 2015, doi: 10.1016/j.jmb.2015.07.016.

[23] A. Fout, J. Byrd, B. Shariat, and A. Ben-Hur, "Protein interface prediction using graph convolutional networks," in Conference on Neural Information Processing Systems, 2017, vol. 30, no. Nips, pp. 6531–6540.

[24] S. Lu, Y. Li, F. Wang, X. Nan, and S. Zhang, "Leveraging Sequential and Spatial Neighbors Information by Using CNNs Linked With GCNs for Paratope Prediction," IEEE/ACM Trans. Comput. Biol. Bioinforma., pp. 1–1, 2021, doi: 10.1109/TCBB.2021.3083001.

[25] J. Meiler, M. Müller, A. Zeidler, and F. Schmäschke, "Generation and evaluation of dimension-reduced amino acid parameter representations by artificial neural networks," J. Mol. Model., vol. 7, no. 9, pp. 360–369, 2001, doi: 10.1007/s008940100038.

[26] S. F. Altschul et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," Nucleic Acids Res., vol. 25, no. 17, p. 33893402, 1997.

[27] S. McGinnis and T. L. Madden, "BLAST: At the core of a powerful and diverse set of sequence analysis tools,"

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452188; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nucleic Acids Res., vol. 32, no. WEB SERVER ISS., pp. 20–25, 2004, doi: 10.1093/nar/gkh435.

[28] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features," Biopolymers, vol. 22, no. 12, pp. 2577–2637, 1983, doi: 10.1002/bip.360221211.

[29] M. F. Sanner, A. J. Olson, and J. Spehner, "Reduced surface: An efficient way to compute molecular surfaces," Biopolymers, vol. 38, no. 3, pp. 305–320, 1996, doi: 10.1002/(sici)1097-0282(199603)38:3¡305::aid-bip4¿3.3.co;2-8.

[30] J. Mihel, M. Šikić, S. Tomić, B. Jeren, and K. Vlahoviček, "PSAIA - Protein structure and interaction analyzer," BMC Struct. Biol., vol. 8, pp. 1–11, 2008, doi: 10.1186/1472-6807-8-21.

[31] P. J. A. Cock et al., "Biopython: Freely available Python tools for computational molecular biology and bioinformatics," Bioinformatics, vol. 25, no. 11, pp. 1422–1423, 2009, doi: 10.1093/bioinformatics/btp163.

[32] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., vol. 9, pp. 1735–1780, 1997.

[33] L. Chen et al., "TransformerCPI: Improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments," Bioinformatics, vol. 36, no. 16, pp. 4406–4414, 2020, doi: 10.1093/bioinformatics/btaa524.

[34] U. C. Berkeley, J. Meier, T. Sercu, and A. Rives, "Transformer Protein Language Models Are Unsupervised Structure Learners," bioRxiv, no. i, pp. 1–24, 2020.

[35] L. A. Staeheli and D. Mitchell, "The Relationship Between Precision-Recall and ROC Curves Jesse," in International Conference on Machine Learning, 2006, pp. 233–240, doi: 10.1145/1143844.1143874.