SEQUENCE-BASED DEEP LEARNING ANTIBODY DESIGN FOR IN SILICO ANTIBODY AFFINITY MATURATION

Yue Kang, Dawei Leng, Jinjiang Guo, ⓘ Lurong Pan*
AIDD Group
Global Health Drug Discovery Institute, Beijing, China
yue.kang@ghddi.org and lurong.pan@ghddi.org

Februrary 17, 2021

ABSTRACT

Antibody therapeutics has been extensively studied in drug discovery and development within the past decades. One increasingly popular focus in the antibody discovery pipeline is the optimization step for therapeutic leads. Both traditional methods and in silico approaches aim to generate candidates with high binding affinity against specific target antigens. Traditional in vitro approaches use hybridoma or phage display for candidate selection, and surface plasmon resonance (SPR) for evaluation, while in silico computational approaches aim to reduce the high cost and improve efficiency by incorporating mathematical algorithms and computational processing power in the design process. In the present study, we investigated different graph-based designs for depicting antibody-antigen interactions in terms of antibody affinity prediction using deep learning techniques. While other in silico computations require experimentally determined crystal structures, our study took interest in the capability of sequence-based models for in silico antibody maturation. Our preliminary studies achieved satisfying prediction accuracy on binding affinities comparing to conventional approaches and other deep learning approaches. To further study the antibody-antigen binding specificity, and to simulate the optimization process in real-world scenario, we introduced pairwise prediction strategy. We performed analysis based on both baseline and pairwise prediction results. The resulting prediction and efficiency prove the feasibility and computational efficiency of sequence-based method to be adapted as a scalable industry practice.

Keywords Graph Neural Networks · Antibody Maturation

1 Introduction

The interaction between antibody and antigen is achieved through the non-covalent bonds between amino acid residues on the interacting surfaces, i.e. the interacting contacts (ICs). A high binding affinity reflects a stable antibody-antigen complex [12], which requires particular amino acids composition as well as conformation upon binding. Researchers have previously reported that at 80 % of the binding free energy is clustered on a very limited number of interacting residues[14]. Specifically, the complementarity-determining region on the antibody, or CDR, is considered to be what allow antibody to bind to a target with high specificity [10, 11, 12]. Although residues on non-interacting surface do not bind with antigen directly, studies have demonstrated that non-interacting surface (NIS) also plays an important role in binding affinity by long-range electrostatic contributions and surface–solvent interactions [9, 13]

Traditional in-silico approaches compute the antibody-antigen binding strength based on atom chemical properties such as polarity and charge[9]. With the fast development in the field of machine learning and deep learning research, algorithms and techniques are explored towards the application of antibody development. [5] studied the prediction of interacting contacts and performed binding prediction using surface-based geometric features; [4] employed topology-based network tree for binding affinity change prediction based on complexes' 3D structures. Efforts to improve antibody-antigen modeling also include deep learning analysis using sequence data only. [3] trained long-short-term memory models on in silico sequence library for antigen-specific affinity predictions. In [8], mutations within CDR-H3

region are modeled and optimized with respect to different antigens by employing ensembled CNN models.In the present study, we proposed GNN-based modeling strategies for antibody maturation by utilizing sequence data only. Base on the Hag-Net graph neural network structure[1], we investigated the contributions of interacting contacts (ICs), non-interacting surface (NIS), as well as antibody-based features in terms of affinity prediction through different modeling designs. The proposed approaches achieved satisfactory classification and regression performance in both five-fold cross-validations and out-of-distribution analysis, and outperformed conventional in-silico analysis on AB-Bind dataset.

2 Data and Methodology

2.1 Data Collection

Antibody-Bind (AB-Bind) is a manually curated and organized database that includes 1101 mutants across 32 complexes, with experimentally determined binding affinities upon mutations [22]. The binding affinities for mutated variants are represented by the change in free energy ($\Delta\Delta G$) of binding in kal/mol. The study also present binging affinity predictions using the bASA, Rosetta, dDFIRE, DFIRE, STATIUM, FoldX, and Discovery Studio scoring functions[22]. The presented methods were further evaluated in terms of their binary classification performance, where mutated variants are categorized as improved vs weakened binders. In the presented study, we follow the same protocols and construct both regression and classification tasks based on AB-Bind dataset. For more details on AB-Bind data and reported computations, the reader is referred to [22].

2.2 Methodology

2.2.1 Hag-Net network structure

The graph neural networks [2] are neural network architectures that designed specifically to cope with graph data. Nodes in the graph are designed to learn an embedding containing information about their associated neighborhoods. The embeddings can be utilized as characteristic features for node labeling, edge prediction as well as graph representation with proper readout/pooling methods. Traditional machine learning applications use preprocessing techniques to map graph structure data into simpler representations, whereas the graph neural networks extends existing neural network methods in order to retains the important information such as the topological dependency of information on each node . The intrinsic design of GNN makes it well-suited to study the molecular and biological interactions and properties, we therefore leverage the power of graph representations to improve the antibody-antigen interaction modeling in the present study.

Specifically, we employed the Hag-Net graph neural network for antibody affinity change prediction. The Hag-Net network structure is invented and pioneered by the AIDD Group [1] of Global Health Drug Discovery Institute. As a variant of spatial based graph neural networks, Hag-Net aggregates features by taking both summation heterogeneous graph based deep learning for biomedical network link prediction and maxima of neighbouring features [1], and therefore increase effect information propagation from shallow layers to deep layers. For graph G = (V, E) with $v, u \in V$ and $e_{uv} \in E$, the generic layer formulation with heterogeneous aggregations is as:

$$h_v = \psi(C(h_v, \bigoplus_{i=1}^{M-1} \phi_i(A_{i,u \in N(v)}(\{(h_v, h_u, e_{uv})\}))))) \tag{1}$$

Eq(1) describes the update method for center node h_v. A represents M different aggregation operators, while $\bigoplus$ represents the merge operator for M neighborhood aggregation results. $C()$ updates h_v with the merged aggregation result. ψ and ϕ_i are linear/non-linear transform functions.

In this application, we utilized the graph-representation of Hag-Net for label prediction,which requires the READOUT function in order to aggregate node features to form the graph-level output:

$$h_G = \psi(\bigoplus_{i=1}^{M-1} \phi_i(A_{i,v \in V}(\{(h_v)\}))) \tag{2}$$

The final classifier layer (consists of several dense layers) takes the graph-level output as input and outputs model prediction for assigned tasks.
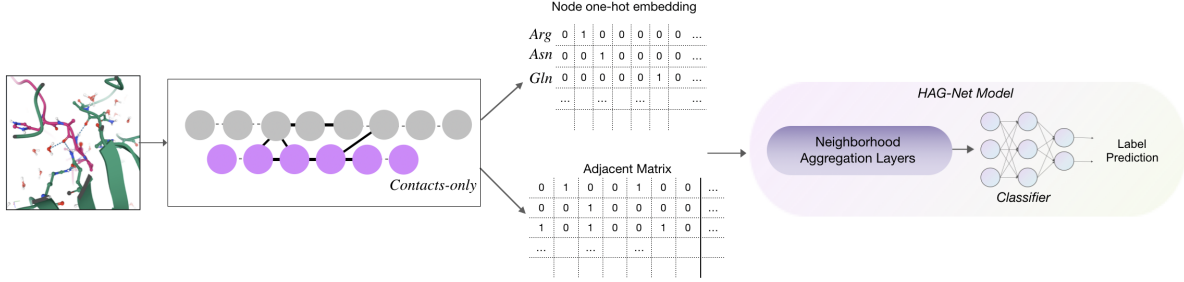
Figure 1: Contacts-only modeling for weakened vs improved binder classification based on Hag-Net network structure.

2.2.2 Baseline study: *binding affinity change ($\Delta\Delta G$) prediction*

We performed antibody-antigen complex affinity prediction based on sequence data with Hag-Net network structure. In the proposed modeling approach, each node represents a single residue, and its edges represent connections with associated/interacting residues. This modeling approach represents each antibody-antigen complex as a single graph structure that corresponds to the amino acid sequences of antibody and antigen. The resulting graph was then converted to an input matrix by one-hot encoding, where each row represents a specific residue, as well as an adjacent matrix, where each 1 represents connection between amino acids in two positions. Thus, a 200 amino acid sequence will result in a $200x22$ matrix as the node input and a $200x200$ adjacent matrix as edge input. The prediction labels were generated as in [22], where binary classification was performed on improved vs weakened binders, and regression prediction was performed on the free binding energy changes values. Figure 1 describes the modeling workflow for binary classification task.

Following the above described workflow, we explore the graph representation of antibody-antigen complex with three different representation strategies. 1) Full-seq model: the full-seq model simply takes antibody and antigen sequences as two separated graph sequences(Figure 2(A)). The intuition is to incorporate both interact contacts (ICs) and non-interacting surface(NIS) into modeling as the binding strength between antibody and antigen relies on the full conformation of the formed complex [9, 12] . 2) Contacts-only model: the contacts-only model produces a compact representation of the complex by utilizing residues on the interfacial surface (distance <5 Angstrom) only(Figure 2(B)); Given limited high-quality training data, this approach is presumably more adequate since it models the most relevant information for antibody-antigen interactions. The identification of interface contacts was obtained using Prodigy-based prediction service[9] based on complex's 3D structure. 3) Antibody-only model: the antibody-only model aims to address the promiscuous binding capability of antibodies cross diverse antigens [15, 16, 17, 18]. We investigated on antibody-only modeling for binding affinity prediction, and evaluate if Hag-Net based network structure captures the enabling features for antibodies' natural binding capability.

2.2.3 Pairwise-study: *binding affinity pairwise rank prediction*

Antibody maturation aims to optimize the binding affinity based on known antibody leads targeting specific antigens. To this purpose, we construct the pairwise problem and study the binary relations between mutated variants of each therapeutic lead. Specifically, for each wildtype antibody, its associated mutations are grouped into unique ordered pairs. Each pair consists of two complexes with same target antigen but different mutated variants. If the first complex ranks higher than the second complex with respect to their binding affinities (i.e. the first complex possesses higher affinity), the pair is labeled as 1, otherwise as 0. Therefore, our goal is to obtain classifier f :

$$f(a,b) = \begin{cases} 1 & \text{if } \Delta\Delta G_a < \Delta\Delta G_b \\ 0 & \text{else} \end{cases}$$

Note that only mutations of the same therapeutic lead are ranked and grouped into ordered pairs. Our pairwise task is designed to predict the binary relations between antibody mutations, in order to evaluate if Hag-Net based model captures the enabling features that contribute to the affinity improvement upon mutations. By modeling the pairwise classification labels instead of binding free energy value, this approach also alleviates the possible errors and biases in the curated data due to different data sources and assays. Figure 3 illustrates the prediction workflow based on Hag-Net modeling for a single ordered pair of antibody complex 1BJ1.
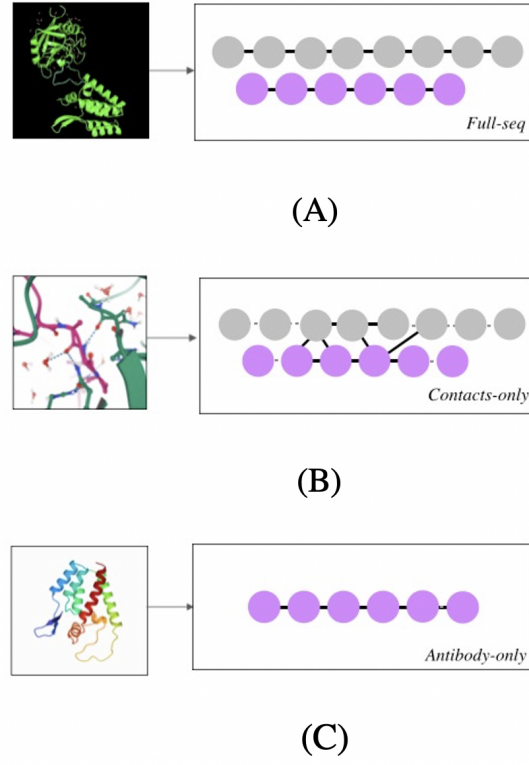
Figure 2: Graph representations of antibody-antigen interaction modeling. (A) Full-seq graph (B) Contacts-only graph (C) Antibody-only graph
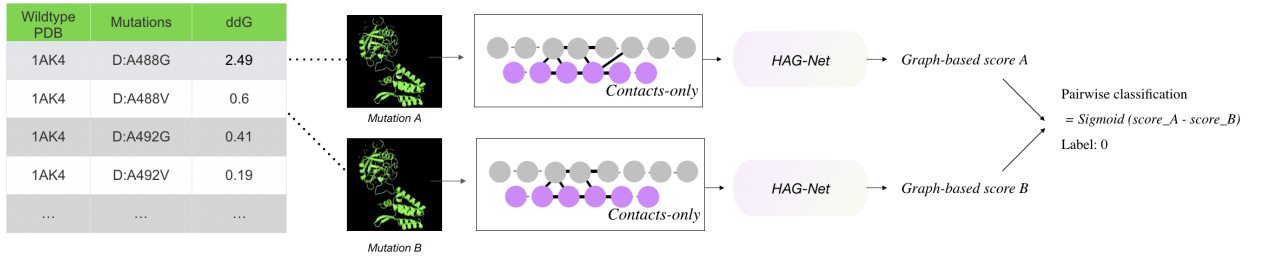

Figure 3: Pairwise classification model with contacts-only representation. In the presented table, each wildtype (e.g. 1AK4) are listed with their associated mutated and binding affinity measurement. To construct an ordered pair, we take two mutations (mutation A and mutation B) and generate corresponding graph representations respectively. The outputs of Hag-Net are used as affinity scores for the comparison between mutations. The difference between affinity scores then goes through sigmoid function to predict the binary relation label, specifically in this example, label is set to 0 since mutate A has lower binding affinity than mutate B.
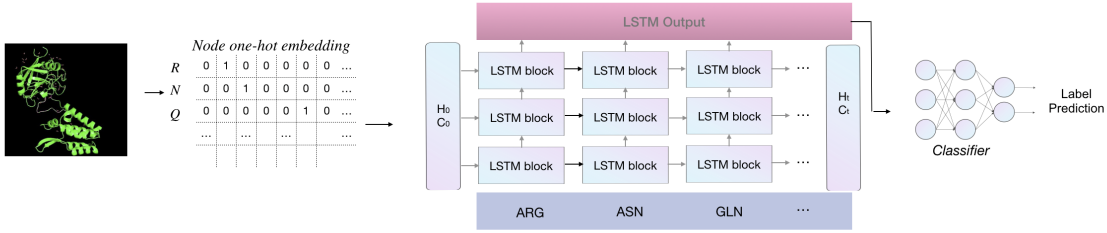
Figure 4: LSTM-based antibody-only classification modeling. Sequence information is converted to an input matrix by one-hot encoding and feed into a three-layer LSTM network.

2.2.4 Benchmark study: *LSTM vs Hag-Net*

Finally, we compared the Hag-Net based network structure with the long short-term memory (LSTM) model, the latter was extensively used in natural language processing field and was previously examined in the research for protein sequence analysis [Mason et al]. Both models were trained using the same set of training data to assess their performance and scalability. Figure 4 illustrates the LSTM approach for affinity prediction task.

3 Results

We investigated through different modeling approaches for in silico antibody maturation optimization . Both five-folds and out-of-distribution cross-validation were performed to evaluate model capability as well as scalability. In the out-of-distribution experiments, five antigens were randomly selected from 32 antigens with their associated antibody-antigen complexes defined as the testing set (i.e. leave-five-out) , whereas the rest 28 antigens and associated complexes were used as the training set. Performance of classification and regression were assessed by constructing ROC area under curve (AUC) and Pearson correlation coefficient respectively.

3.1 Baseline study

Table 1 shows the averaged AUC of binary classification on improved vs weakened binders using Hag-Net based models. Follow the protocol in [22], the testing set was split into subsets based on their $\Delta\Delta G$ absolute values ($|\Delta\Delta G|$>1, $|\Delta\Delta G|$>0.5 and $|\Delta\Delta G|$>0) and evaluated separately (Table 1). Benchmark studies performance from [22]are also listed for comparison. As shown in Table1, the three Hag-Net modeling approaches are able to delineate improved and weakened binders well with respect to all subsets (Table1 (A)); In five-folds cross-validations, Hag-Net models yield satisfactory prediction accuracy in all subsets, and reveals high delineating capability on subsets with $|\Delta\Delta G|$>1 (with AUC > 0.9) ;

The capability of predicting 'unseen' complex structures is assessed by comparing the benchmark performance (Table1(A)) with out-of-distribution performance based on Hag-Net (Table1(B)). Specifically, for $|\Delta\Delta G|$>0, Hag-Net based approaches yield comparable delineation accuracy when comparing to the best performance reported (AUC = 0.73 from Discovery Studio); For samples with more significant changes in their binding affinities ($|\Delta\Delta G|$>0.5 and $|\Delta\Delta G|$>1), Hag-Net significantly outperforms conventional approaches. This demonstrates that Hag-Net based approaches can be used in predictions on completely unseen variants during antibody maturation. We hypothesized that Hag-Net captures the transferable features that contributes to the binding strength in antibody-antigen interactions.

We observed that the antibody-only modeling, when comparing to the other two approaches which incorporated antigen into modeling, shows solid delineating capability for weakened vs improved binders classifications, in both five-folds validations and novel variants predictions. This delineating capability observed in deep learning models is in line with antibodies naturally binding capability across antigens as previously reported [15, 16, 17, 18]

Table 2 shows the regression performance using the metric of Pearson correlation coefficients. Full-seq and contacts-only approaches yield satisfactory regression performance in five-folds validations. $\Delta\Delta G$ value prediction on unseen variants are less significant yet still comparable with the best performance among conventional in silico approaches. We noticed that the antibody-only models possess less accuracy in both five-folds and out-of-distribution experiments, as it does not capture antibody-antigen specificity into modeling.

Table 1. Baseline Classification: Improved binder vs Weakened binder

(A) Model Results and Performance Comparision : five-folds cross-validation

Methods	|ddG| > 0	|ddG| > 0.5	|ddG| > 1.0	
STATIUM	0.64 (0.60 - 0.67)	0.74 (0.70 - 0.78)	0.81 (0.76 - 0.85)	AUC values are listed with the 95 percent confidence intervals
FoldX	0.70 (0.67 - 0.73)	0.81 (0.78 - 0.85)	0.87 (0.83 - 0.91)	
Discovery Studio	0.73 (0.70 - 0.76)	0.82 (0.78 - 0.86)	0.88 (0.85 - 0.92)	
HagNet_Full-Seq	**0.86(0.02)**	**0.97(0.01)**	**0.99(0.004)**	Averaged AUC values are listed with std value among 5-folds studies
HagNet_Contacts-only	0.85(0.02)	0.96(0.02)	**0.99(0.007)**	
HagNet_Antibody-only	0.84(0.01)	**0.97(0.01)**	**0.99(0.002)**	

(B)Model Results and Performance Comparision : leave-five-out (out-of-distribution performance)

Methods	|ddG| > 0	|ddG| > 0.5	|ddG| > 1.0	
HagNet_Full-Seq	0.71(0.03)	0.84(0.06)	0.92(0.07)	Averaged AUC values are listed with std value among 5-folds studies
HagNet_Contacts-only	**0.74(0.07)**	0.85(0.04)	**0.96(0.01)**	
HagNet_Antibody-only	0.70(0.05)	**0.88(0.04)**	0.95(0.05)	

Table 2. Baseline comparison: Regression task

Methods		Five folds	Out-of-distribution	
STATIUM	0.32			
FoldX	0.34			
Discovery Studio	0.45			
HagNet_Full-Seq		0.70	**0.45**	
HagNet_Contacts		**0.72**	0.43	
HagNet_Antibody-Seq		0.66	0.39	

3.2 Pairwise Study

Overall, 97711 pairs were generated from the original 1101 mutants of 32 complexes. The positive/negative ratio is 1 due to the intrinsic design for ordered pairs (i.e. for each positive pair, a negative pair was generated). We split the data into training and testing set (80/20). Table 3 shows the pairwise task performance (AUC) with respect to full-seq, antibody-only and contacts-only representations. The pairwise problem aims to find favorable mutated variants of the same therapeutic lead with respect to their binding affinity, as opposed to baseline studies, in which different wildtypes and associated mutations are broadly defined as either improved or weakened binders. In essence, we anticipate the pairwise study to predict the proper ranks of the antigen-specific binding strength on mutated variants.

Note that as the proposed pairwise study does not consider the affinities between different therapeutic leads (i.e. wildtypes), its results are not applicable to be compared with baseline studies. $|\Delta\Delta G|$ subset does not apply in pairwise study either due to the intrinsic design of ordered pair generation. The averaged pairwise classification performance are presented in Table 3.

In Table 3, all three proposed graph representations achieved excellent results in five-folds studies, suggesting that Hag-Net based modeling is feasible to be used for the antibody maturation process. Specifically, full-seq modeling significantly outperforms the other two approaches in out-of-distribution experiments with AUC 0.70, indicating that both interacting contacts (ICs) and non-interacting surface (NIS) contribute to the affinity changes on therapeutic leads upon mutations.

Table 3. Pairwise study classification performance

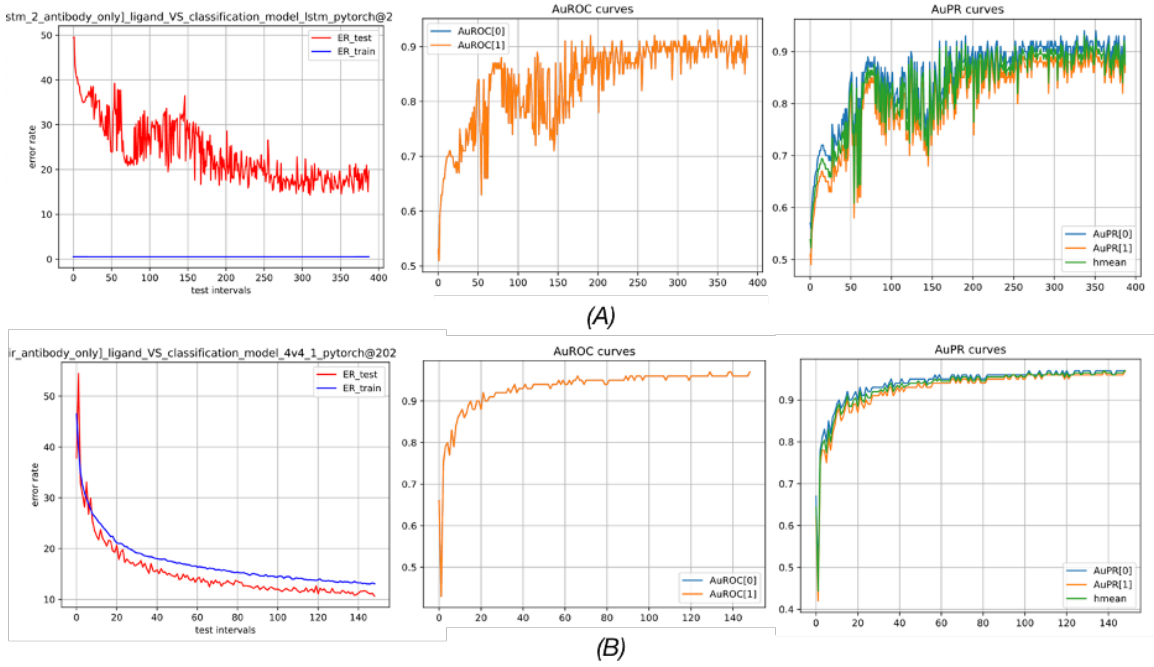| Methods	Five-folds	Out-of-distribution
HagNet_Full-Seq	**0.96**	**0.70**
HagNet_Contacts	0.96	0.65
HagNet_Antibody-Seq	0.95	0.65

Figure 5: Convergence curves of error rate(left), AUC (middle) and AUPR(right) for class 1 and class 0 on pairwise dataset (A) LSTM-based model (B) Hag-Net based model

3.3 Deep Learning Benchmark Study

To further evaluate the Hag-Net adequateness as the backbone structure for antibody-antigen interaction modeling, we incorporate the long short-term memory (LSTM) based modeling as an alternative approach and compare the two strategies with respect to training efficiency, prediction accuracy along and generalization capability. Antibody-only representation was employed for model simplicity. Figure 5 depicts the convergence curves of LSTM based and Hag-Net based training on one of five folds during cross-validations in pairwise study.

In Figure 5, we observed that Hag-Net converges faster than LSTM during training with higher prediction accuracy (performance showed LSTM at epoch 400 and Hag-Net at epoch 150), and results in more stable and consistent training performance. Table 3 demonstrates the prediction accuracy of the two approaches , where Hag-Net model yields better delineation in five-folds analysis and lesser performance in out-of-distribution tasks.

Table 4. Pairwise study comparison LSTM vs Hag-Net

Methods	Five-folds	Out of distribution
LSTM_Antibody-only	0.89	**0.68**
HagNet_Antibody-only	**0.95**	0.65

4   Discussions

In the last several decades, the number of known protein sequences has grown exponentially along with the development of sequencing technologies, yet the number of corresponding crystal structures is limited due to their high dependency on experiments. To examine the potential of sequence data- guided in silico antibody maturation, we proposed affinity prediction models that utilize deep learning techniques based on the complex sequence information only. The intuition of such modeling strategy is to capture the enabling features encrypted at amino acids level that contributes to the interactions and the resulting binding strength between antibody and antigen. By modeling at the level of amino acids level instead of atoms, the proposed approach significantly reduces the time cost for both training and prediction process, while possesses better tolerance on the size of annotated training datasets due to the model's simplicity. This structure-free optimization approach also allows it applicability on broader scope of analysis given the inevitable disparity between available sequence data and crystal structures in antibody and protein research field.

Since the discovery of antibodies, researcher have been searching for the interpretation of relationship between limited number of antibodies and their binding ability to the almost infinite antigens space [15]. Despite antibodies' exquisite specificity to target antigens, studies have reported and examined antibodies promiscuous binding capabilities [15]. [18] postulated that the presence of antibodies' poly-reactivity may be caused by the flexible antigen-binding 'pocket', which can change conformation to accommodate different antigens. Such cross-reactivity is usually studied as a challenge for high specificity antibody development [17], while on the other hand it also allows the invention on broad-specificity antibody thru engineering approaches [16]. In our antigen-independent experiments, antibody-only models have exhibited solid prediction capability for the delineations between improved vs weakened binders, and can be further applied on unseen and novel samples outside the training distribution. This might suggest that such the 'natural binding' capability possessed by antibodies can be modeled by graph-based network structure with proper model designs, and ought to be further investigated for antibody maturation development with stronger data diversity and larger sample size.

With the propagation between graph-convolutional layers and pyramid features intrinsically designed in the graph-based structure, the well-trained Hag-Net models are expected to represent the antibody-antigen complex in a much more comprehensive fashion comparing to traditional machine learning approaches. This is demonstrated by the higher accuracy (AUC 0.95 vs AUC 0.89) and significantly faster and smoother training achieved in Hag-Net when comparing to LSTM models (Figure 5). We notice, however, that the Hag-Net model inevitably suffers from larger number of parameters due to its structure and is therefore inclined to the overfitting problem given limited training data, resulting in the lesser generalization performance in out-of-distribution tasks.

Through our analysis, the prediction performance and model scalability are inevitably limited by the size of publicly available dataset with annotated binding affinity. The proposed method and trained models ought to be validated on larger dataset, and its feasibility in assisting the antibody maturation need to be further evaluated along with industry practice on unseen wildtype samples. We are also aware that the proposed modeling for antibody-antigen interactions does not guarantee its applicability in other protein-protein interactions (PPI), as antibodies possess unique properties and does not fall into the general protein interaction space.

5   Conclusion

We proposed Hag-Net based deep learning modeling approaches for antibody maturation based on sequence data. The contributions on interacting contacts (ICs), non-interacting surface (NIS) in terms of affinity prediction are studied through different modeling designs, along with the antibodies' naturally binding capability. The proposed approaches achieved excellent performance in both five-fold cross-validations and out-of-distribution analysis, and outperformed conventional in-silico analysis on AB-BIND dataset. Our pairwise study further investigated Hag-Net models' feasibility in studying the antigen-specific binding strength under the therapeutic leads optimization scenario. In the end, we performed benchmark comparison by using the LSTM-based backbone structure for modeling, assessing training efficiency as well as model performance between language modeling and graph-based modeling in the application of antibody-antigen analysis. We concluded that the proposed methodology, though limited by sample size, possesses high efficiency and robust delineation capability for antibody affinity prediction and therefore holds great potential to be further validated and utilized in the antibody development pipeline [1].

6   Acknowledgement

References

[1] Leng, Dawei, et al. "Enhance Information Propagation for Graph Neural Network by Heterogeneous Aggregations." arXiv preprint arXiv:2102.04064 (2021).

[2] Scarselli, Franco, et al. "The graph neural network model." IEEE transactions on neural networks 20.1 (2008): 61-80.

[3] Mason, Derek M., et al. "Deep learning enables therapeutic antibody optimization in mammalian cells by deciphering high-dimensional protein sequence space." BioRxiv (2019): 617860.

[4] Wang, Menglun, Zixuan Cang, and Guo-Wei Wei. "A topology-based network tree for the prediction of protein–protein binding affinity changes following mutation." Nature Machine Intelligence 2.2 (2020): 116-123.

[5] Gainza, Pablo, et al. "Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning." Nature Methods 17.2 (2020): 184-192.

[6] Graves, Jordan, et al. "A review of deep learning methods for antibodies." Antibodies 9.2 (2020): 12

[7] Balci, A. T., et al. "DeepInterface: Protein-protein interface validation using 3D Convolutional Neural Networks." bioRxiv (2019): 617506.

[8] Liu, Ge, et al. "Antibody complementarity determining region design using high-capacity machine learning." Bioinformatics 36.7 (2020): 2126-2133.

[9] Xue, Li C., et al. "PRODIGY: a web server for predicting the binding affinity of protein–protein complexes." Bioinformatics 32.23 (2016): 3676-3678.

[10] Maccallum, R.M.; Martin, A.C.R.; Thornton, J.M. Antibody-antigen Interactions: Contact Analysis and Binding Site Topography. J. Mol. Biol. 1996, 262, 732–745.

[11] Sela-Culang, Inbal, Vered Kunik, and Yanay Ofran. "The structural basis of antibody-antigen recognition." Frontiers in immunology 4 (2013): 302.

[12] Dondelinger, Mathieu, et al. "Understanding the significance and implications of antibody numbering and antigen-binding surface/residue definition." Frontiers in immunology 9 (2018): 2278.

[13] Kastritis, Panagiotis L., et al. "Proteins feel more than they see: fine-tuning of binding affinity by properties of the non-interacting surface." Journal of molecular biology 426.14 (2014): 2632-2652

[14] Robin G, Sato Y, Desplancq D, Rochel N, Weiss E, Martineau P. Restricted diversity of antigen binding residues of antibodies revealed by computational alanine scanning of 227 antibody–antigen complexes. J Mol Biol. (2014) 426:3729–3743. doi: 10.1016/j.jmb.2014.08.013

[15] James, Leo C., Pietro Roversi, and Dan S. Tawfik. "Antibody multispecificity mediated by conformational diversity." Science 299.5611 (2003): 1362-1367.

[16] Leivo, Janne, et al. "Engineering of a broad-specificity antibody: detection of eight fluoroquinolone antibiotics simultaneously." Analytical biochemistry 409.1 (2011): 14-21.

[17] Buchwalow, Igor, et al. "Non-specific binding of antibodies in immunohistochemistry: fallacies and facts." Scientific reports 1.1 (2011): 1-6.

[18] Notkins, Abner Louis. "Polyreactivity of antibody molecules." Trends in immunology 25.4 (2004): 174-179.

[19] Gamble, Theresa R., et al. "Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid." Cell 87.7 (1996): 1285-1294.

[20] Sehnal, D. R. A. S., et al. "Mol*: towards a common library and tools for web molecular graphics." Workshop on Molecular Graphics and Visual Analysis of Molecular Data. The Eurographics Association. Vol. 10. 2018.

[21] DeLano, Warren Lyford. "PyMOL." (2002): 700.

[22] Sirin, Sarah, et al. "AB-Bind: antibody binding mutational database for computational affinity predictions." Protein Science 25.2 (2016): 393-409.