



## OPEN ACCESS

## EDITED BY

Dirk Werling,  
Royal Veterinary College (RVC),  
United Kingdom

## REVIEWED BY

Prasanna Srinivasan Ramalingam,  
Vellore Institute of Technology, India  
Dong Xia,  
Royal Veterinary College (RVC),  
United Kingdom

## \*CORRESPONDENCE

Dongsup Kim  
✉ kds@kaist.ac.kr

RECEIVED 14 June 2025

ACCEPTED 18 July 2025

PUBLISHED 13 August 2025

## CITATION

Choi S and Kim D (2025) BIDpred: unraveling  
B cell Immunodominance hierarchical pattern  
using statistical feature discovery and deep  
learning prediction.  
*Front. Immunol.* 16:1646946.  
doi: 10.3389/fimmu.2025.1646946

## COPYRIGHT

© 2025 Choi and Kim. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC BY\)](#).  
The use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# BIDpred: unraveling B cell Immunodominance hierarchical pattern using statistical feature discovery and deep learning prediction

Sungjin Choi and Dongsup Kim\*

Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology,  
Daejeon, Republic of Korea

Knowledge of B cell immunodominance is important for designing vaccines that may elicit effective immune responses. However, the prevalence and characteristics of B cell immunodominance remain poorly understood. In this study, we introduced an immunodominance score through novel data processing methods and identified statistically significant characteristics of B cell immunodominance at the residue and patch levels. Based on these findings, we developed BIDpred, a B cell ImmunoDominance predictor, that learns newly discovered features by leveraging protein language model embeddings and graph attention network to predict the immunodominance scores. BIDpred demonstrates superior performance in predicting immunodominance scores compared to existing methods while maintaining competitive accuracy with state-of-the-art methods for conventional B cell epitope prediction. To the best of our knowledge, this is the first study to systematically analyze and predict B cell immunodominance patterns, marking a significant advancement in vaccine design research.

## KEYWORDS

immunoinformatics, B cell immunodominance, vaccine design, deep learning, protein language model

## Introduction

Immunodominance (ID), described as a hierarchical level of preference for immune response in antigen, is essential for understanding adaptive immunity. In vaccine design, ID is often used to elicit the intended immune response in epitope-based vaccines (1). While T cell ID has been extensively studied (2–4), relatively little is known about B cell ID (5). Furthermore, even the existence of B cell epitopes has been argued that epitopes could be everywhere on the surface of antigens (6). Despite this ambiguity, structural studies of antibody-antigen complexes, such as those involving influenza A and SARS-CoV-2, reveal

certain levels of immunological preference among antigen amino acid residues; some residues of antigens appear more likely to bind antibodies (7, 8).

Numerous computational approaches have been developed to predict B cell epitopes for vaccine design applications (9–21). However, previous methods have overlooked the hierarchical nature of ID between epitopes as they were solely trained on the binary labels by classifying residues as either epitopes or non-epitopes. These labels were typically defined using distance-based annotations derived from antigen-antibody complex structures. We conjecture that incorporating ID scores, estimated using a novel data processing strategy, could provide additional insights into the relative immunological preference of certain epitopes over others. ID scores are continuous scores ranging from 0 to 1, which represent the epitope priorities against the antibody interaction. Although the ID score defined in this study may not perfectly capture the biological phenomenon due to limited data availability, it could still serve as a reasonable approximation to enhance our understanding of epitope prioritization.

We defined the ID score through a comprehensive data curation process, which involves sequence clustering, multiple sequence alignment (MSA) building, and epitope annotation. Statistical analyses were then conducted to investigate the physico-chemical, geometrical, evolutionary, and compositional characteristics associated with B cell ID. Residue- and patch-level analyses revealed statistically significant correlations with several features,

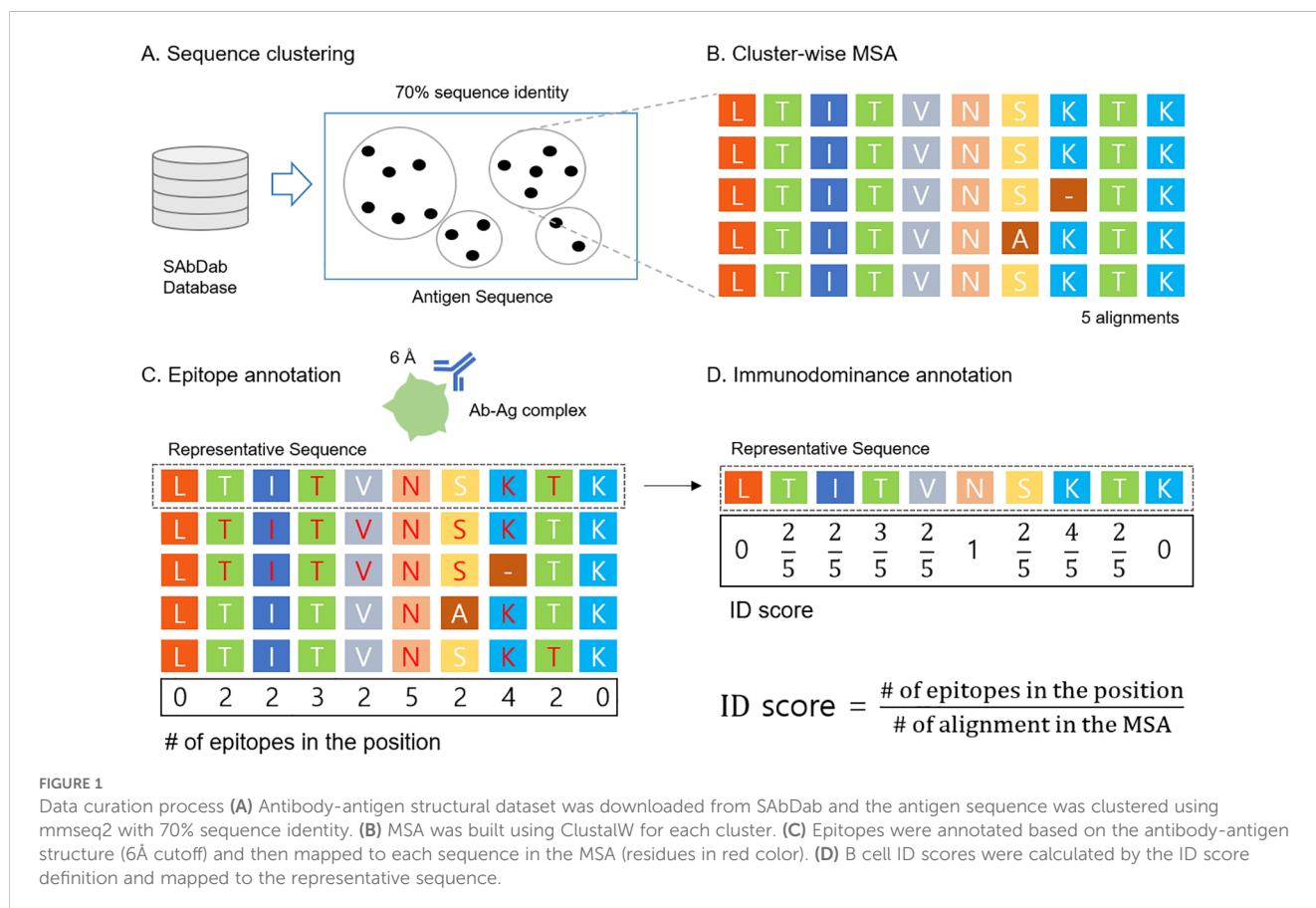
particularly strong signals related to conservation and clustering patterns.

Building on these insights, we developed BIDpred, a predictive model that leverages highly significant features identified in the statistical analysis. BIDpred integrates pretrained protein language model embeddings with a graph attention network (GAT) to capture the nuanced features of B cell ID. BIDpred demonstrates superior performance in predicting immunodominance scores compared to existing methods while maintaining comparable accuracy with the state-of-the-art methods for conventional B cell epitope prediction. This highlights the value of our novel data processing approach in providing additional insights into immune response preferences among epitopes. Code and dataset are available at <https://github.com/sj584/BIDpred> and webserver at <http://bidpred.kaist.ac.kr>.

## Materials and methods

### Data curation

X-ray crystallography data were collected from SAbDab (22) as of Mar.19, 2024. The data curation process is illustrated in Figure 1. For quality filtering, we used a filtering cutoff of resolution 3.0Å, R-factor 0.25, antigen size with at least 50 amino acids, and maximum antibody sequence identity of 99% (23). After quality filtering, we



extracted the antigen sequences and performed sequence clustering using the `mmseq2 easy-cluster` command with a minimum sequence identity threshold of 0.70; all other options were set to default (24) (Figure 1A). Clusters with at least 4 elements were used. One cluster can be viewed as a group of the same antigens interacting with different antibodies. Within each cluster, MSAs were constructed using the Clustal Omega `clustalo` command with default settings (25) (Figure 1B). Further information about the MSA depth for each cluster can be found in Supplementary Table S1. In each MSA, we annotated epitopes from the antibody-antigen complex structures. Antigen residues within 6 Å distance to antibody residues were defined as epitopes. After epitope annotation was mapped to each sequence in MSA (Figure 1C), ID scores were assigned to the representative sequence of the MSA (Figure 1D). In summary, each data point corresponds to a representative protein from its own cluster, with the ID score annotated based on epitope annotation in cluster-wise MSA.

Motivated by the previous study (26), ID score is defined by,

$$ID\ score = \frac{\# \text{ of epitopes in the position}}{\# \text{ of alignment in the MSA}}$$

Since the number of epitopes at a given position in an MSA is always less than or equal to the number of alignments in the MSA, ID score could only range from 0 to 1. For the data split, we reasoned that the representative sequence with more alignments in the MSA would better reflect the characteristics of B cell ID. Consequently, we split the data such that the test set included representative sequences with at least 10 alignments in the MSA, while the training sets contained representative sequences with 4 to 9 alignments in the MSA. This resulted in a total of 92 training sets and 24 test sets. The distribution of the antigen type can be found in Supplementary Figure S1 and a full list of antigen types is provided in Supplementary Table S11. Other characteristics such as antigen size, epitope, non-epitope, and immunodominance can be found in Supplementary Figure S2.

## Feature embeddings

A molecular graph was generated from the protein structure. Nodes were amino acid residues, and edges were constructed when nodes were close to each other within 10 Å distance. ESM-based residue representation was used for node embedding. For ESM-2 (27), `esm2_t33_650M_UR50D` model was used. For ESM-IF1 (28), `esm_if1_gvp4_t16_142M_UR50` model was used.

## Statistical analysis of protein features

Statistical analysis was performed on the representative sequences in the test set, which has representative sequences with at least 10 alignments in MSA. We used an independent sample T-test for amino acid feature analysis, while a Mann-Whitney U-test was used for compositional analysis. The null hypothesis was rejected at a significance level of  $\alpha \leq 0.05$ . Afterwards, we

performed Benjamini-Hochberg procedure for multiple testing correction. Geometric features, physico-chemical features, compositional (either amino acid or secondary structure), and evolutionary features were examined at both residue and patch levels (Supplementary Table S2). Surface residues with at least relative surface accessibility (RSA) 0.10 were used. Surface patches were constructed by grouping surface residues within a 10 Å distance of central surface residues.

## Features for statistical analysis

RSA and secondary structure were collected from DSSP (29) module in Biopython (30). Protrusion and residue depth were obtained from PSAIA (31). Hydrophobicity, isoelectric point, residue volume, steric, polarizability, H-bond donor, polarity, positive charge, and negative charge were acquired from AAIndex (32) (Supplementary Table S3). Per-residue conservation score was obtained from ConSurf (33).

## Hyperparameters

Hyperparameters were optimized to achieve the lowest validation loss. Specifically, 200 epochs, batch size of 4, 8 multi-attention heads, learning rate of 1e-6, Adam optimizer, mean squared error loss, 3 GAT layers with hidden dimensions of 2048-512-128, and 2 fully connected (FC) layers with dimensions of 128-32-1 were used.

## Evaluation metrics

B cell ID is a hierarchical level of preference in immune response. Therefore, Spearman correlation was the main criterion for measuring the ID. Also, R-squared ( $R^2$ ) score and Pearson correlation were used for additional evaluation. For B cell epitope prediction, area under the receiver operating characteristic curve (AUC-ROC) and area under the precision-recall curve (AUC-PR) were used for threshold independent evaluation for classification.

## Result

### Statistical analysis reveals B cell ID in residue and patch level

To investigate which characteristics are related to the B cell immunodominant regions, statistical analysis was performed to examine geometrical, physicochemical, compositional, and evolutionary features at both the residue and patch levels (Supplementary Table S2). Only surface residues with a relative surface accessibility (RSA) of at least 0.10 were considered. Surface patches were generated from the central node with the same surface criteria. We defined the two ID groups based on the following criteria;

$$\begin{cases} \text{highly immunodominant, if } ID \geq 0.5 \times \text{max immunodominance of protein} \\ \text{weakly immunodominant, if } 0 < ID < 0.2 \times \text{max immunodominance of protein} \end{cases}$$

In residue-wise analysis (Figure 2, Supplementary Table S4), highly immunodominant residues exhibit distinct patterns in several features, including residue volume, polarizability, hydrogen bond donor, and conservation. Specifically, highly immunodominant residues tend to have bigger residue volume, attractive interactions mediated by electrons, act as stronger hydrogen bond donors, and display greater variability in sequence conservation.

In patch-level analysis (Figure 3, Supplementary Table S4), we observed that features identified in residue-level analysis are still valid, implying that amino acids with similar properties are located close to each other. We additionally found different patterns in steric, RSA, and protrusion with statistical significance. In the patch level, highly immunodominant patches were additionally found to be highly steric, less exposed, and less protruding.

Based on the observation that similar residues are close to each other in patch analysis, we investigated the distribution of immunodominant residues within a patch by analyzing neighbor immunodominance. For this analysis, residue ID scores in the patch

were averaged while excluding the central node ID score (Figure 4A). Here, a drastic difference pattern was observed (Figure 4B, Supplementary Table S4). The result demonstrates immunodominant residues tend to cluster closely together. This clustering pattern was further visualized using PyMol (34), which clearly highlights the spatial grouping of immunodominant residues in B cell ID (Figures 4C–E).

## B cell ID prediction task

Thus far, we have identified several features associated with B cell immunodominance (ID). With the aforementioned features in mind, we further explored B cell ID prediction methods using deep learning. However, most prediction models based on statistical features failed to achieve satisfactory performance, likely due to weak signals that were insufficiently detectable. Meanwhile, we noted that conservation and ID clustering patterns showed strong statistical significance with low p-value (Supplementary Table S4). Thus, we devised a model which leverages evolutionary and clustering features in an efficient way (Figure 5). We used ESM-2 (27) sequence embedding and ESM-IF1 (28) structure embedding to capture the evolutionary features of the B cell ID. To capture the

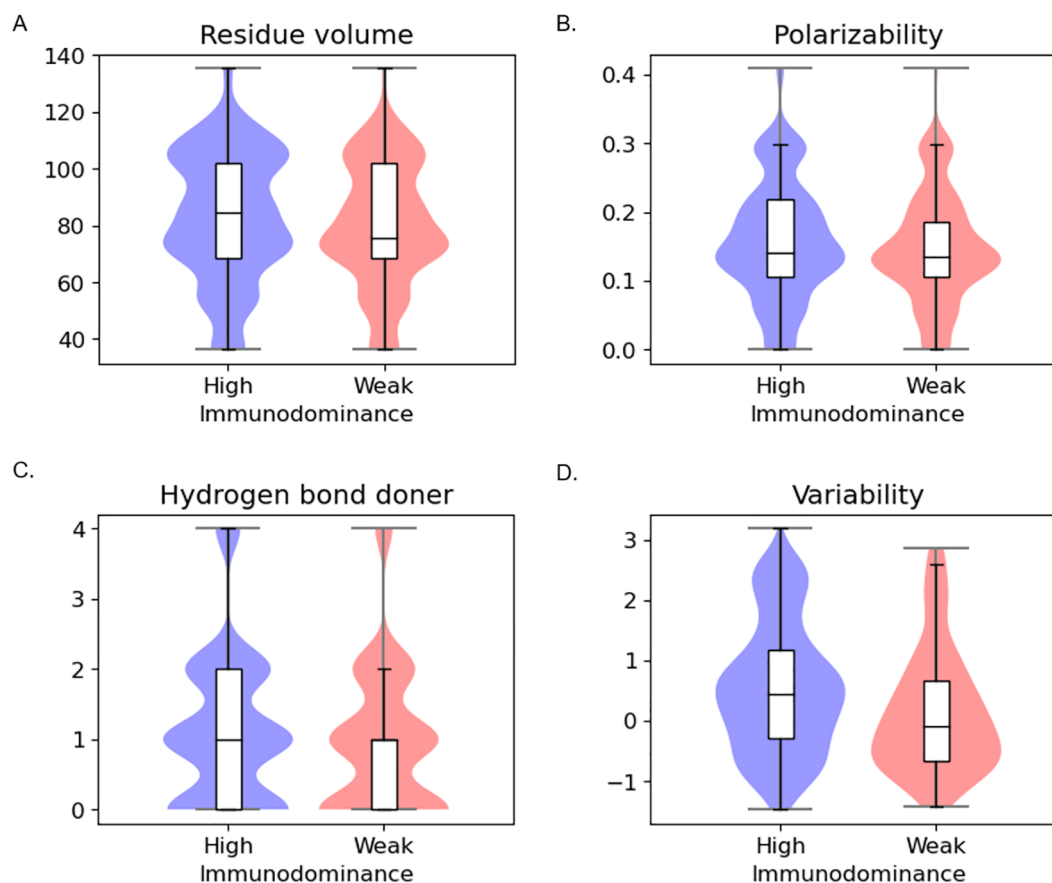


FIGURE 2

Residue-level statistical analysis between high immunodominance and weak immunodominance groups (A) Residue volume (B) Polarizability (C) Hydrogen bond donor (D) Conservation.

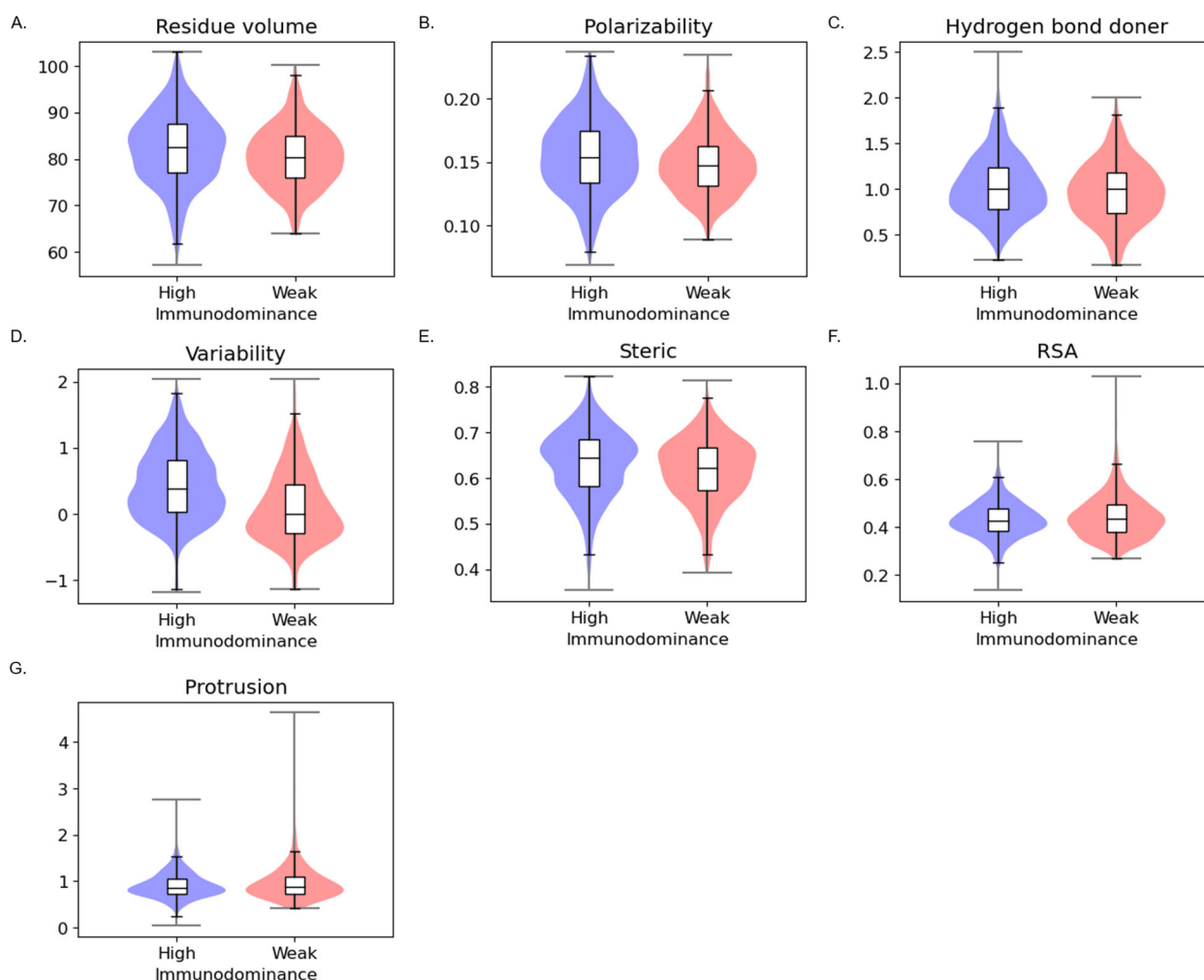


FIGURE 3

Patch-level statistical analysis between the high immunodominant and weak immunodominant groups (A) Residue volume (B) Polarizability (C) Hydrogen bond donor (D) Conservation (E) Steric (F) RSA (G) Protrusion.

structural homophily of the B cell ID, we used graph attention network which uses message passing node updates from the adjacent nodes.

We evaluated our model on B cell ID benchmark datasets and compared it with existing B cell epitope prediction methods (Table 1). The main criterion is Spearman correlation coefficient, which measures the hierarchical preference of the immunogenic residues. For this analysis, redundancy was removed by ensuring 70% sequence identity between each test set and the training sets of the B epitope predictors. Since each model has a different training set, subset of our test set was used for evaluation in comparison with our model. This resulted in four separate evaluations. Across 24 test sets, our model consistently achieved the highest Spearman correlation scores, outperforming all other models. While ElliPro, SEPPA3, and CBTope showed almost random performance in ID prediction, BepiPred-3 and DiscoTope-3 were reported to have certain levels of capturing immunodominance (Supplementary Table S5). We reasoned that those models, being trained on ESM-based evolutionary features with their redundant training sets,

might have partially learned ID patterns. Evaluation results for all models using independent and the same test set are provided in Supplementary Table S6.

## BIDpred method showed comparable results in conventional B cell epitope prediction

We also compared our model with other methods in the conventional B cell epitope prediction task (Table 2). We used Epitepe3D (35) benchmark dataset with 45 PDB. Again, the redundancy in the test set was removed by applying a 70% sequence identity threshold. In this evaluation, our model demonstrated results comparable to those of current state-of-the-art models. Notably, this performance was achieved despite our model being exclusively trained on the B cell ID prediction task. This outcome supports our initial hypothesis that the B cell ID prediction task captures essential features relevant to B cell epitope prediction.

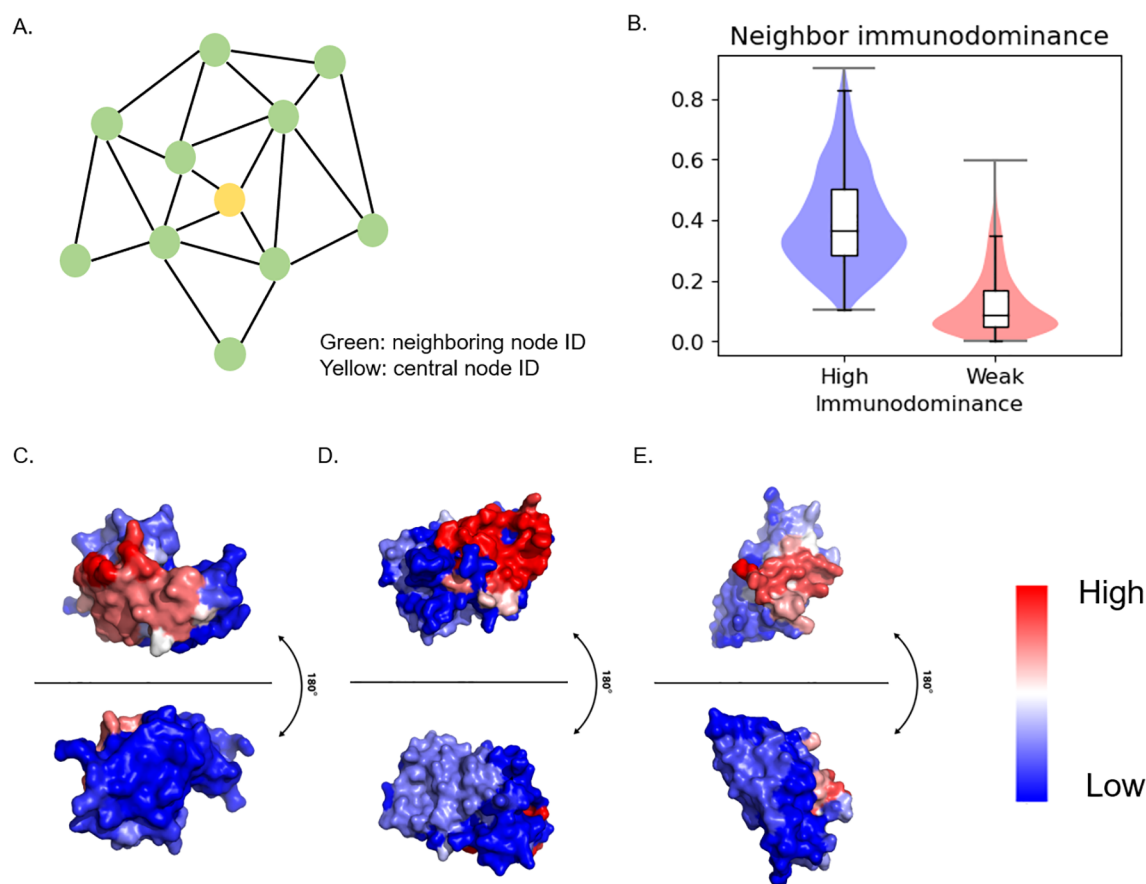


FIGURE 4

Neighboring immunodominance analysis and visualization **(A)** Schematics of Surface patch. Patch was generated from 10 Å distance from the central node. Patch ID was defined by central node ID. Neighbor ID was the average value of the neighboring node ID score. **(B)** Statistical analysis of neighbor immunodominance. **(C)** PDB ID: 1FBI, X chain **(D)** PDB ID: 4YPG, D chain **(E)** PDB ID: 8JEL, J chain.

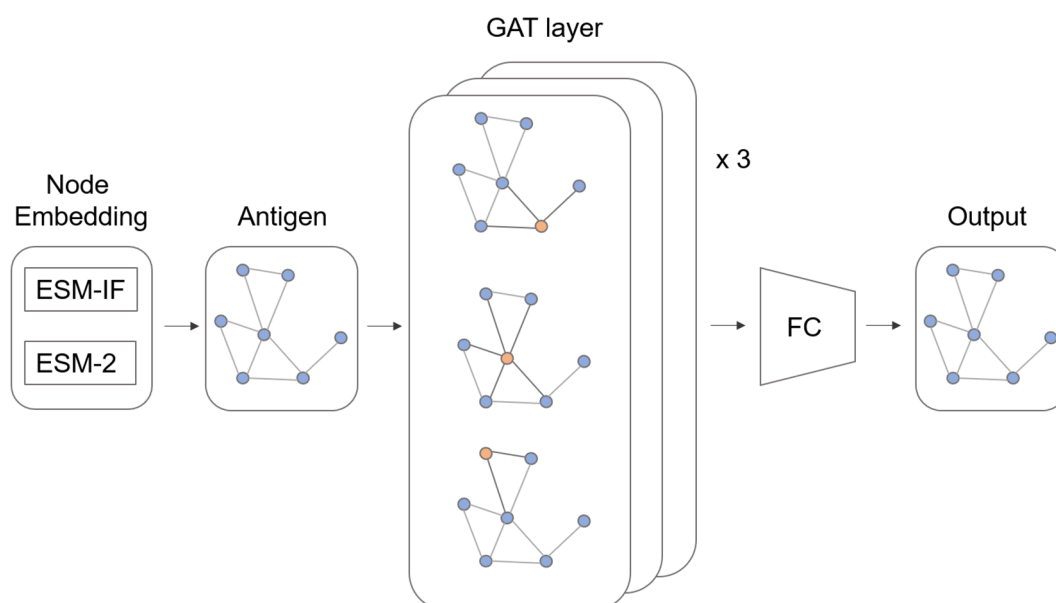


FIGURE 5

Model architecture. ESM-based pretrained model embeddings were used for node features. The antigen structure was represented as a molecular graph and encoded by graph attention (GAT) layer. A fully connected (FC) layer predicts the immunodominance score.



**TABLE 1** B cell ID benchmark in comparison to conventional B cell epitope predictors.

| Method      | Pearson      | Spearman     | R2 score     | # of PDB test |
|-------------|--------------|--------------|--------------|---------------|
| ElliPro     | 0.078        | 0.057        | -5.165       |               |
| Ours        | <b>0.393</b> | <b>0.406</b> | <b>0.144</b> | 24/24         |
| SEPPA3      | 0.088        | -0.015       | -0.096       |               |
| Ours        | <b>0.422</b> | <b>0.396</b> | <b>0.169</b> | 14/24         |
| CBTope      | 0.065        | 0.017        | -1.250       |               |
| Ours        | <b>0.374</b> | <b>0.346</b> | <b>0.136</b> | 20/24         |
| Bepipred-3  | 0.338        | 0.437        | 0.096        |               |
| DiscoTope-3 | 0.405        | 0.444        | 0.018        |               |
| Ours        | <b>0.429</b> | <b>0.476</b> | <b>0.181</b> | 6/24          |

Bold values indicate the highest score in the benchmark.

## Ablation study

We conducted an ablation study to evaluate the impact of the features used in our ID prediction model (Table 3, Supplementary Table S7). The study revealed that both ESM features outperformed the conventional one-hot encoding method, commonly used as a feature representation. Among the two, ESM-IF1 (structural evolutionary features) proved to be more effective than ESM-2 (sequential evolutionary features). The best performance was achieved when both features were combined, highlighting their complementary nature in capturing B cell ID characteristics. The superior importance of structural evolutionary features can be attributed to the spatial clustering of highly immunodominant residues in 3D space, emphasizing the role of structural context in B cell ID learning.

## Case study: SARS-CoV-2

We performed a case study on SARS-CoV-2 in the test set (Figure 6), which included a large number of structures, resulting in the most extensive MSA alignments (271 alignments). In Figure 6A,

**TABLE 2** B cell epitope prediction benchmark from the epitope3D external dataset.

| Method        | AUC  | AUC-PR | MCC   |
|---------------|------|--------|-------|
| SEPPA 3       | 0.55 | 0.11   | -0.02 |
| ElliPro       | 0.48 | 0.09   | -0.02 |
| epitope3D     | 0.49 | 0.08   | -0.03 |
| CBTope        | 0.55 | 0.13   | 0.01  |
| Bepipred-3.0  | 0.71 | 0.20   | 0.08  |
| DiscoTope-3.0 | 0.71 | 0.19   | 0.09  |
| Ours          | 0.71 | 0.21   | 0.15  |

**TABLE 3** Ablation study on ID prediction test set.

| Ablation | Pearson      | Spearman     | R2 score     |
|----------|--------------|--------------|--------------|
| One-hot  | -0.027       | -0.006       | -0.052       |
| ESM-2    | 0.328        | 0.302        | 0.097        |
| ESM-IF1  | 0.367        | 0.368        | 0.123        |
| Ours     | <b>0.393</b> | <b>0.406</b> | <b>0.144</b> |

Bold values indicate the highest score in the benchmark.

we visualized the ID pattern and prediction results along the residue positions. The model predictions closely aligned with the SARS-CoV-2 ID pattern, accurately identifying ID peaks in most cases. This agreement was reflected in strong Spearman, Pearson, and  $R^2$  scores. Additionally, the scatter plot shown in Figure 6B illustrates a clear linear relationship between the model predictions and the actual ID pattern, further validating the model's predictive accuracy for SARS-CoV-2. In comparison with other previous tools, our model demonstrated the best performance, as shown in Supplementary Table S8.

## Discussion

B cell ID remains relatively unexplored from statistical and computational perspectives, despite its biological significance. For example, in the influenza A virus, the hemagglutinin (HA) head is typically the primary target of immune responses, whereas the HA stem is a subdominant region. However, the HA head is highly mutable, which often renders previous vaccines ineffective against new variants. To address this challenge, vaccine development efforts are focusing on targeting the less immunodominant but more conserved HA stem, aiming to create universal vaccines (36, 37). In our dataset, influenza A virus ID scores are high on the HA head instead of the stem. Also, the statistical analysis on variability aligns with the conservation pattern in Influenza A virus, which suggests that our data processing well approximates the biological pattern. For many pathogens with unclear ID patterns due to limited data, ID prediction tools can serve as a valuable resource for vaccine discovery, facilitating the identification of conserved and strategically targetable regions.

In summary, we systematically investigated B cell immunodominance (ID) through statistical analysis and developed a prediction model tailored to this task. We curated a dataset specifically designed to capture B cell ID using a variety of bioinformatics tools. Our analysis identified key characteristics associated with B cell ID, highlighting statistically significant features. After statistically identifying the B cell ID, we developed a deep learning method, BIDpred, specifically trained on the B cell ID dataset. We proposed that the conventional B cell epitope prediction task closely aligns with B cell ID prediction task, but existing models were not optimally trained for the B cell ID prediction. BIDpred demonstrated the ability to perform both tasks effectively by capturing the essence of antibody-agnostic B cell epitope prediction. To our knowledge, this work represents the first

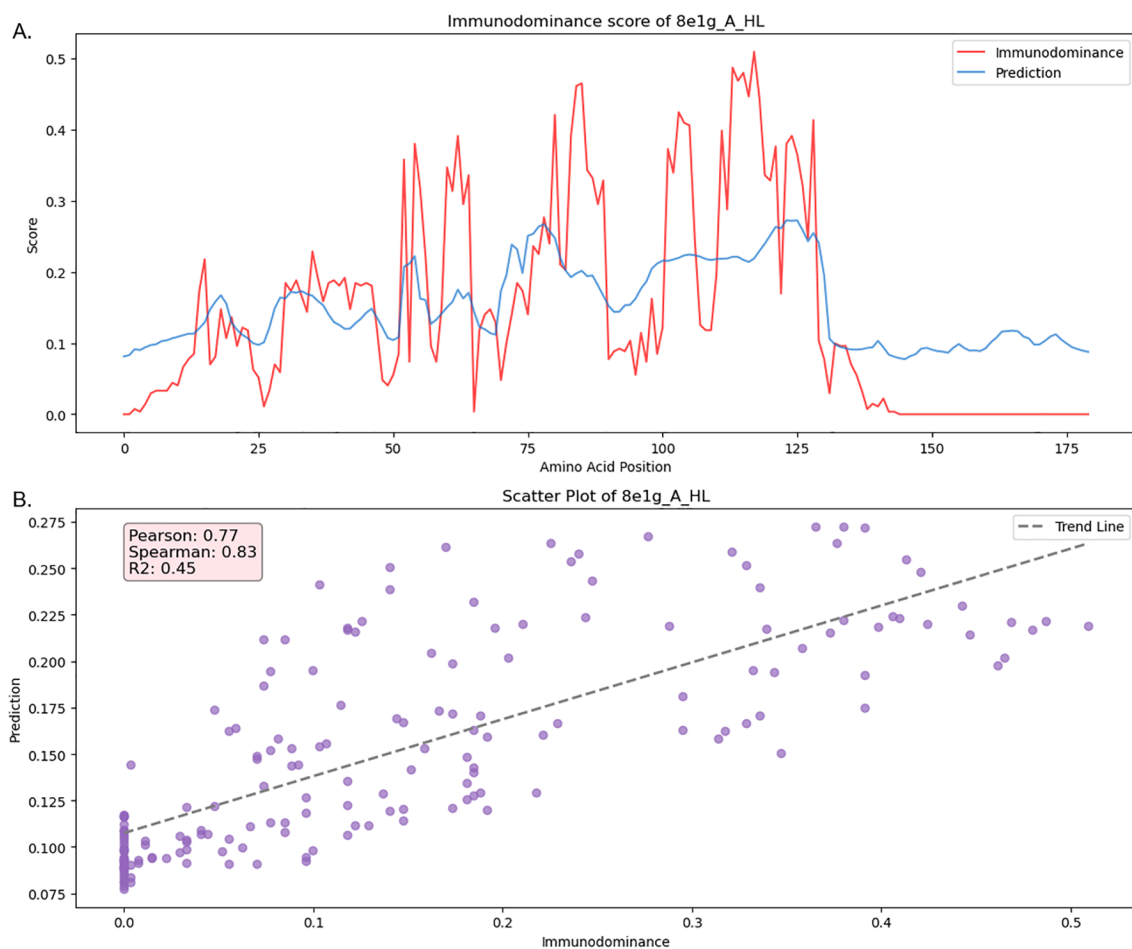


FIGURE 6

Case study of SARS-CoV-2 immunodominance pattern and model prediction from test set (PDB ID: 8e1g) (A) Immunodominance/prediction score per amino acid position (B) Scatter plot of prediction and immunodominance with evaluation metrics.

attempt to statistically analyze B cell ID and to train a deep learning model optimally for B cell ID prediction, paving the way for advancements in vaccine design and immunological research.

## Conclusion

B cell ID is crucial for vaccine design. However, few studies about B cell ID were available. In this study, we introduced a B cell ID score, conducted a comprehensive statistical analysis of B cell ID features, and developed a deep learning-based prediction model. Our findings and predictive tools have the potential to accelerate vaccine development and stimulate further research into B cell ID, enhancing our understanding of adaptive immunity.

The limitation of this work is the deficiency of the data being clustered, which might lead to incomplete B cell ID annotation. Each data point is a representative protein of all proteins in a single cluster. While the representative proteins in the test set were selected to have at least 10 alignments in the MSA, those in the training set had fewer alignments, ranging from 4 to 9. This discrepancy suggests that the training set may not fully capture the robustness of the test set,

potentially affecting model performance. In [Supplementary Table S9](#), we present results from random shuffling of the train and test sets. When the test set was randomly selected, it often contained very few representative proteins with high MSA depth, resulting in noised training and test outcomes. However, when we ensured that the test set included at least half of its representative proteins with high MSA depth, model training and testing became more stable. This suggests that the rigorous evaluation of model predictions requires a test set with a sufficient number of representative proteins with high MSA depth. As more antigen-antibody structural data becomes available, the dataset could better approximate the true B cell ID phenomenon, leading to enhanced model training and improved predictive accuracy.

There are other limitations regarding the gap between computational biology and experimental biology. While our statistical analysis shows correlation of immunodominance, these findings do not necessarily indicate biological causation. Therefore, further experimental validation is required before these results can be translated into biological application. Additionally, some of the characteristics identified as significant, such as RSA and Protrusion in patch level, are correlated with each other, drawing attention to the result in terms of redundancy. ([Supplementary Figure S3](#)) While our



model generalizes well to several types of antigens including virus, parasite, and human proteins, (Supplementary Table S10) the model has a limitation of not distinguishing neutralizing epitopes from non-neutralizing epitopes. However, it is reasonable to expect that the epitopes with high ID scores imply the neutralizing epitopes. To our knowledge, no prediction tools have attempted to predict direct neutralization. Future direction should aim for predicting the neutralizing effect to bridge the gap between physical interaction and biological function to enhance translational relevance.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material. Code and data are available at <https://github.com/sj584/BIDpred> and webserver is available at <https://bidpred.kaist.ac.kr/>.

## Author contributions

SC: Software, Investigation, Data curation, Formal analysis, Writing – review & editing, Visualization, Validation, Methodology, Writing – original draft. DK: Supervision, Writing – review & editing, Resources, Conceptualization, Funding acquisition, Project administration.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported

by National Research Foundation of Korea grants funded by Korean Government [grant numbers RS-2024-00344154, RS-2024-00399520, RS-2025-00523107].

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2025.1646946/full#supplementary-material>

## References

- Musunuri S, Weidenbacher PA, Kim PS. Bringing immunofocusing into focus. *NPJ Vaccines*. (2024) 9:11. doi: 10.1038/s41541-023-00792-x
- Akram A, Inman RD. Immunodominance: a pivotal principle in host response to viral infections. *Clin Immunol*. (2012) 143:99–115. doi: 10.1016/j.clim.2012.01.015
- Schreiber H, Wu T, Nachman J, Kast W eds. *Immunodominance and tumor escape*. *Seminars in cancer biology*. Elsevier (2002).
- Kim PS, Lee PP, Levy D. A theory of immunodominance and adaptive regulation. *Bull Math Biol*. (2011) 73:1645–65. doi: 10.1007/s11538-010-9585-5
- Dale GA, Shartouny JR, Jacob J. Quantifying the shifting landscape of B cell immunodominance. *Nat Immunol*. (2017) 18:367–8. doi: 10.1038/ni.3695
- Kunik V, Ofra Y. The indistinguishability of epitopes from protein surface is explained by the distinct binding preferences of each of the six antigen-binding loops. *Protein Eng Des Sel*. (2013) 26:599–609. doi: 10.1093/protein/gzt027
- Angeletti D, Gibbs JS, Angel M, Kosik I, Hickman HD, Frank GM, et al. Defining B cell immunodominance to viruses. *Nat Immunol*. (2017) 18:456–63. doi: 10.1038/ni.3680
- Dugan HL, Stamper CT, Li L, Changrob S, Asby NW, Halfmann PJ, et al. Profiling B cell immunodominance after SARS-CoV-2 infection reveals antibody evolution to non-neutralizing viral targets. *Immunity*. (2021) 54:1290–303. e7. doi: 10.1016/j.immuni.2021.05.001
- Clifford JN, Høie MH, Deleuran S, Peters B, Nielsen M, Marcotili P. BepiPred-3.0: Improved B-cell epitope prediction using protein language models. *Protein Sci*. (2022) 31:e4497. doi: 10.1002/pro.4497
- Høie MH, Gade FS, Johansen JM, Würtzen C, Winther O, Nielsen M, et al. DiscoTope-3.0-Improved B-cell epitope prediction using AlphaFold2 modeling and inverse folding latent representations. *Frontiers in Immunology*. (2024) 15:1322712. doi: 10.3389/fimmu.2024.1322712
- Jespersen MC, Peters B, Nielsen M, Marcotili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*. (2017) 45:W24–W9. doi: 10.1093/nar/gkx346
- Liang S, Zheng D, Standley DM, Yao B, Zacharias M, Zhang C. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinf*. (2010) 11:1–6. doi: 10.1186/1471-2105-11-381
- Ponomarenko J, Bui H-H, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinf*. (2008) 9:1–8. doi: 10.1186/1471-2105-9-514
- Swerdoski MJ, Baldi P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics*. (2008) 24:1459–60. doi: 10.1093/bioinformatics/btn199
- Park M, Seo S-W, Park E, Kim J. EpiBERTope: a sequence-based pre-trained BERT model improves linear and structural epitope prediction by learning long-distance protein interactions effectively. *bioRxiv*. (2022). doi: 10.1101/2022.02.27.481241
- Shashkova TI, Umerenkov D, Salnikov M, Strashnov PV, Konstantinova AV, Lebed I, et al. SEMA: Antigen B-cell conformational epitope prediction using deep transfer learning. *Front Immunol*. (2022) 13:8272. doi: 10.3389/fimmu.2022.960985

17. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol*. (2012) 8:e1002829. doi: 10.1371/journal.pcbi.1002829
18. Collatz M, Mock F, Barth E, Hölzer M, Sachse K, Marz M. EpiDope: a deep neural network for linear B-cell epitope prediction. *Bioinformatics*. (2021) 37:448–55. doi: 10.1093/bioinformatics/btaa773
19. Ansari HR, Raghava GP. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. *Immunome Res*. (2010) 6:1–9. doi: 10.1186/1745-7580-6-6
20. Zhou C, Chen Z, Zhang L, Yan D, Mao T, Tang K, et al. SEPPA 3.0—enhanced spatial epitope prediction enabling glycoprotein antigens. *Nucleic Acids Res*. (2019) 47:W388–W94. doi: 10.1093/nar/gkz413
21. Ivanisenko NV, Shashkova TI, Shevtsov A, Sindeeva M, Umerenkov D, Kardymon O. SEMA 2.0: web-platform for B-cell conformational epitopes prediction using artificial intelligence. *Nucleic Acids Res*. (2024) 52:gkae386. doi: 10.1093/nar/gkae386
22. Dunbar J, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, et al. SABDab: the structural antibody database. *Nucleic Acids Res*. (2014) 42:D1140–D6. doi: 10.1093/nar/gkt1043
23. Krawczyk K, Liu X, Baker T, Shi J, Deane CM. Improving B-cell epitope prediction and its application to global antibody-antigen docking. *Bioinformatics*. (2014) 30:2288–94. doi: 10.1093/bioinformatics/btu190
24. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. (2017) 35:1026–8. doi: 10.1038/nbt.3988
25. Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. *Multiple Sequence Alignment Methods: Springer*. (2013) p:105–16. doi: 10.1007/978-1-62703-646-7\_6
26. Cia G, Pucci F, Rooman M. Critical review of conformational B-cell epitope prediction methods. *Briefings Bioinf*. (2023) 24:bbac567. doi: 10.1093/bib/bbac567
27. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*. (2023) 379:1123–30. doi: 10.1126/science.ade2574
28. Hsu C, Verkuil R, Liu J, Lin Z, Hie B, Sercu T, et al eds. *Learning inverse folding from millions of predicted structures. International Conference on Machine Learning*. Baltimore, Maryland, USA, PMLR (2022) 162, 17–23.
29. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Orig Res Biomol*. (1983) 22:2577–637. doi: 10.1002/bip.360221211
30. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. (2009) 25:1422. doi: 10.1093/bioinformatics/btp163
31. Mihel J, Šikić M, Tomić S, Jeren B, Vlahoviček K. PSAIA—protein structure and interaction analyzer. *BMC Struct Biol*. (2008) 8:1–11. doi: 10.1186/1472-6807-8-21
32. Kawashima S, Kanehisa M. AAindex: amino acid index database. *Nucleic Acids Res*. (2000) 28:374–. doi: 10.1093/nar/28.1.374
33. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. (2016) 44:W344–W50. doi: 10.1093/nar/gkw408
34. Schrödinger L. The PyMOL molecular graphics system. *Version*. (2010) 1:0.
35. da Silva BM, Myung Y, Ascher DB, Pires DE. epitope3D: a machine learning method for conformational B-cell epitope prediction. *Briefings Bioinf*. (2022) 23:bbab423. doi: 10.1093/bib/bbab423
36. Xu D, Carter JJ, Li C, Utz A, Weidenbacher PA, Tang S, et al. Vaccine design via antigen reorientation. *Nat Chem Biol*. (2024), 1–10. doi: 10.1038/s41589-023-01529-6
37. Paules CI, Marston HD, Eisinger RW, Baltimore D, Fauci AS. The pathway to a universal influenza vaccine. *Immunity*. (2017) 47:599–603. doi: 10.1016/j.immuni.2017.09.007