Assessing the Predictive Ability of Computational Epitope Prediction Methods on Fel d 1 Allergen

Running title: Fel d 1 epitope prediction

Jungjoon K. Lee[1,†], Soobon Ko[2,†], Kyungsoo Ha[3], Yoonjoo Choi[4*]

[1] Toolgen, Seoul, Republic of Korea

[2] Department of Agricultural and Biological Chemistry, Chonnam National University, Gwangju, Republic of Korea

[3] New Drug Development Center, Osong Medical Innovation Foundation, Osong, Republic of Korea.

[4] Combinatorial Tumor Immunotherapy MRC, Chonnam National University Medical School, Hwasun-gun, Jeollanam-do

[†] These authors contributed equally.

Correspondence

Yoonjoo Choi, DPhil (Oxon.)

Combinatorial Tumor Immunotherapy MRC

Chonnam National University Medical School

Hwasun-gun, Jeollanam-do 58128, South Korea.

Telephone: +82) 61-379-8477

Fax: +82) 61-379-8476

E-mail: kalicuta@gmail.com

Abstract

Purpose:

Though computational epitope prediction methods have been widely used, there have been only limited studies conducted in the context of allergies. Our research aims to benchmark publicly available epitope prediction tools, focusing on Fel d 1, whose allergenic IgE- and T-cell epitopes have been extensively studied.

Methods:

Our study utilized an array of epitope prediction tools publicly accessible via the Immune Epitope Database (IEDB) and other resources. The tools were evaluated based on their ability to identify known linear IgE- and T-cell epitopes of Fel d 1.

Results:

In general, B-cell epitope prediction methods demonstrated limited effectiveness. Most methods perform marginally better than random selection. ElliPro, a structure-based method, slightly outperformed the rest, suggesting that incorporating 3D structure information could enhance prediction accuracy. In terms of T-cell epitope prediction, ProPred was successful in identifying all known T-cell epitopes, whereas the IEDB approach missed two known epitopes and showed a high rate of over-prediction.

Conclusions:

Our results show that current computational epitope prediction methods possess limitations in accurately identifying allergenic Fel d 1 epitopes. The study highlights the scope for future advancements in computational epitope prediction methodologies and the development of extensive epitope databases to optimize allergenic epitope prediction tools. Despite the evident limitations, these tools can still provide valuable preliminary insights into potential allergenic regions within proteins.

Keywords: Allergenic Epitope, Computational Epitope Prediction, Cat Allergy, Fel d 1

Introduction

The cat allergy is one of the most common pet allergies, affecting approximately 20% of the global population (1). The severity of cat allergy can vary, ranging from mild symptoms like coughing, pruritus, and skin eruption to potentially life-threatening reactions such as anaphylaxis. The primary cause of cat allergy is an allergenic protein called Fel d 1, which is found in the fur and dander of domestic cats. It is estimated that approximately 90% of cat allergies are caused by Fel d 1 (2).

Fel d 1 is primarily secreted by epithelial cells and the salivary glands of cats, and it remains on their haircoats during grooming. One notable characteristic of Fel d 1 is its high thermal stability, allowing it to persist on cat hair and become airborne. It can associate with small airborne particles, spreading throughout the surroundings and leading to allergic reactions in susceptible individuals. In fact, a study reported that Fel d 1 was detected in 99.9% of households in the United States, highlighting its ubiquity and potential for exposure (3).

Although the function of Fel d 1 remains largely unknown, its allergenicity of Fel d 1 has been extensively studied. When Fel d 1 is presented by antigen-presenting cells, such as dendritic cells, it triggers the production of IgE antibodies during sensitization. These IgE antibodies specifically target and bind to epitopes on Fel d 1, leading to the activation of immune cells and the subsequent release of inflammatory mediators. This immune response causes the clinical manifestation of allergic symptoms (4). In addition, T cell epitopes recognized by helper T cells can indirectly induce IgE production through T cell-mediated reactions, further amplifying the allergic response (5). Thus, the recognition of IgE bound to effector cells, such as basophils and mast cells, plays a crucial role in the allergic cascade (4).

Previous studies have revealed that the total IgE reactivity against the natural heterodimer of Fel d 1 is higher compared to its monomeric forms, suggesting that the epitopes of Fel d 1 are partially conformational in nature (6, 7). It has been reported that there are three essential IgE

epitopes in Fel d 1, which are recognized by IgE antibodies produced in allergic individuals (8).

Accurate identification and understanding of epitopes related to allergies are essential for comprehending allergic responses and developing diagnostic tools. Experimental identification of B-cell epitopes has traditionally been the "gold standard" for epitope mapping (9). However, this approach is often costly, time-consuming, and requires specialized laboratory techniques. Alternatively, computational epitope prediction methods have emerged as practical tools for epitope identification (10, 11). These methods utilize algorithms and machine learning techniques to analyze the physicochemical properties of proteins and predict potential epitopes. Early computational tools for B-cell epitope prediction relied on amino acid propensity scales, which characterized the physicochemical properties of B-cell epitopes such as hydropathy (12, 13), flexibility (14), and surface accessibility (15). These scales provided a foundation for initial predictions but had limitations in accuracy and specificity. However, recent advancements in machine learning techniques have leveraged computational B-cell epitope prediction methods, leading to improved outcomes. Machine learning algorithms, including Support Vector Machine (SVM) (16-18), Random Forest (RF) (19, 20), K-nearest neighbors (KNN) (21), and Manifold Adaptive Experimental Design (MAED) (22), have been developed and trained to differentiate known B-cell epitopes from non-B-cell epitopes (17-20, 23-29). These machine learning-based approaches utilize large datasets of experimentally validated epitopes to identify patterns and features that distinguish epitopes from non-epitopes. By learning from this training data, the algorithms can make predictions on new protein sequences and provide insights into potential epitope regions.

While there have been many computational methods for B-cell epitope prediction, their application in predicting IgE epitopes, specifically in the context of allergenic proteins like Fel d 1, has been largely limited. IgE epitopes are of particular interest due to their direct

involvement in allergic reactions. Therefore, in this study, we aimed to evaluate the effectiveness of currently available computational T- and B-cell epitope prediction tools applied to Fel d 1 and their capacity to accurately identify allergenic epitopes.

To accomplish this, we employed an assortment of epitope prediction tools accessible through the Immune Epitope Database (IEDB) (30) and other sources. Our objective was to perform a detailed assessment of these methods, focusing on their ability to identify Fel d 1 allergenic epitopes.

Materials and Methods

Defining Fel d 1 Epitopes

While it is known that some Fel d 1 epitopes are partially conformational, our focus in this study was on the linear epitopes of Fel d 1. This choice was made due to the general low consistency and accuracy of computational prediction methods for conformational epitopes (31).

Fel d 1 is a heterodimer composed of two polypeptide chains (**Fig. 1A**), known as Chain 1 and Chain 2 (also referred to as α and β chains, respectively). Chain 1 consists of 70 amino acids and has an approximate molecular mass of 8 kDa. Chain 2 exhibits variation in its C-terminal region, resulting in three known isoforms (32-34). The major linear IgE-binding epitopes of Fel d 1 are located in three regions: two in Chain 1 and one in Chain 2 (Chain 1: 26 VAQYKALPVVLENA 38, 46 DAKMTEEDKRNALS 59, Chain 2: 12 DVFFAVANGNELLL 25) (35). T-cell epitopes are present in a total of six regions, including overlapping segments (**Table 1**) (36).

Epitope Prediction Tools

To computationally predict Fel d 1 epitopes, we mainly used epitope prediction tools from the

Immune Epitope Database (IEDB). The IEDB hosts a comprehensive suite of seven sequence-based B-cell epitope prediction tools, along with two structure-based prediction methods. For the structure-based methods, we utilized the dimeric structure of Fel d 1 (PDB ID: 1PUO). Among the structure-based tools, we specifically chose ElliPro (37), but excluded DiscoTope (26), as DiscoTope predicted all residues of Fel d 1 to be epitopes. In addition to the IEDB's B-cell epitope prediction tools, we also used BepiPred-3.0 (38), a protein language model epitope predictor. We applied default cut-off values for each prediction method. The following are the methods we benchmarked in this study:

1. BepiPred-1.0, 2.0, and 3.0 (19, 20, 38): Currently, the IEDB webserver offers BepiPred-1.0 and 2.0. BepiPred-1.0 employs a combination of a hidden Markov model (HMM) and a high-performing propensity scale method. BepiPred-2.0 advances further by using a Random Forest (RF) algorithm trained on epitopes annotated from antibody-antigen protein structures. The latest version, BepiPred-3.0, leverages protein language models (LMs) on extensive datasets of protein sequences and structures.
2. Chou and Fashman β-turn prediction (39, 40): A method developed to predict β-turns, based on the idea that it can also assist in predicting antibody epitopes.
3. Emini surface accessibility scale (15): Based on a structural database, it predicts the probability of amino acid being present on the surface from 6mer peptide sequences.
4. Karplus and Schulz flexibility prediction scale (14): It predicts the flexibility of a specific position of an amino acid from the crystal structure's B-factor, thus predicting the probability of being found on the surface.
5. Kolaskar and Tongaonkar antigenicity scale (41): It calculates the tendency of antigenic portions by dividing the hydrophilicity, surface accessibility, and flexibility into 7mer amino acid peptide sequences and calculating the average.
6. Parker hydrophilicity scale (42): An index indicating the degree of peptide

hydrophilicity, calculated by dividing 7mer peptide sequences to estimate amino acid hydrophilicity. The method is based on the idea that areas with high hydrophilicity are likely to be on the surface and therefore become epitopes.

7. ElliPro (37): A structure-based epitope prediction tool. It assumes the protein's surface as an ellipsoid and predicts the protruding areas in relation to the ellipsoid. These areas are then clustered into a protrusion index (PI) to predict the epitopes. Although ElliPro can predict both linear and conformational epitopes, only the linear epitope prediction results were used in this study.

For the prediction of T-cell epitopes, we employed two pMHC-II binding prediction tools: ProPred (43) and the IEDB MHC-II binding prediction tool (44). We set a 5% threshold for ProPred to characterize each nonamer as either a binder or a non-binder for 8 HLA types (45). As for the IEDB prediction method, the query sequence was fragmented into overlapping 15-mer linear peptides, and binding prediction was performed for each peptide using the 27 HLA reference set (46) to ensure broad coverage across the human population. In this study, we used the IEDB recommended 2.22 method, using the percentile rank as a binding indicator. Peptides with a rank below 10 were classified as binders, while those with a higher rank were considered non-binders. The epitope scores for each method were calculated as a linear sum of the number of binding events. To evaluate the known T-cell epitopes shorter than 15 residues using the IEDB prediction tool, extra residues at each terminus were added to make the peptides within a 15mer window (**Table 1**).

Results

Structure Analysis of Fel d 1 Epitopes

Fel d 1 primarily adopts the alpha-helical heterodimeric structure (**Fig. 1A**). The heterodimer is composed of two distinct chains (Chain 1 and 2) that are linked by three disulfide bonds.

These heterodimers naturally assemble into tetramers via non-covalent interactions between Chain 2s, as shown in **Fig. 1B** (47).

Given the structure and formation of the tetramer, the IgE epitope located on Chain 2 might be less accessible for IgE binding when the protein is in its tetrameric structure. As the binding of IgE antibodies to their specific epitopes is a crucial step in the onset of allergic reactions, the non-covalent interactions between Chain 2s in the tetramer might render these epitopes less exposed, thereby decreasing the likelihood of IgE binding.

Prediction of Fel d 1 IgE Epitopes

To assess the IgE epitope prediction, we examined nine publicly available epitope prediction tools, which include web-based B-cell epitope prediction methods provided by the IEDB. Generally speaking, most epitope prediction methods seem to excel more at predicting non-epitopes than epitopes (**Table 2**). However, considering the imbalance between the number of epitopes and non-epitopes, with 42 out of 162 total residues being epitopes (28 out of 70 in Chain 1, and 14 out of 92 in Chain 2), the likelihood of accurately predicting epitopes by mere chance is considerably lower (0.26) compared to that of correctly predicting non-epitopes (0.74). It is important to note that the performance of most methods approximates to random chance. Indeed, some methods like BepiPred-1.0, the Chou β-turn, Parker hydrophilicity, and Karplus flexibility scales even underperform compared to random selection. While the Emini accessibility and Kolaskar antigenicity scales perform marginally better than random chance, their slight improvement could arguably be within the margin of error, making them essentially equivalent to random predictions.

Examining the performance of BepiPred-3.0 more closely, we find that while it achieves the highest precision in non-epitope prediction, its recall rate is extremely low. This suggests that the default cut-off value might be set too low, leading to a high rate of misprediction for non-

epitopes. In a side-by-side comparison of BepiPred-2.0 and ElliPro, a closer inspection of the epitope prediction results by chain suggests that ElliPro outperforms BepiPred-2.0 (**Fig. 2**). This relative superior performance by ElliPro is likely due to its unique approach of using a protein's structure as an input, allowing it to predict epitopes from surface residues, thus giving it a significant edge over other methods.

Prediction of Fel d 1 T-cell Epitopes

T-cell immunity plays a pivotal role in managing allergic responses by dictating the magnitude and nature of the immune response to allergens. In particular, T-cell responses against allergens can have profound influence on the production of specific types of antibodies, such as IgE, which plays a direct role in allergic reactions.

In our initial investigations, we subjected the known Fel d 1 T-cell epitopes to ProPred and the IEDB prediction methods. Among these epitopes, the sequence 55 ENALSLLDKIYTS 67 in Chain 1, known for its binding to HLA-DRB1*14:01, was unable to be evaluated since the specific allele is not included in either of the methods. Notably, we found that the IEDB method was not able to correctly predict two of the known T-cell epitopes as binders for their respective alleles. Contrarily, ProPred correctly identified all the epitopes as binders for their associated alleles (**Table 1**).

It is important to note that high epitope binding scores do not necessarily indicate immunogenicity. Though antigen presentation takes place, actual immune responses are initiated and regulated by T-cell recognition. This leads to the conclusion that the lack of pMHC binding almost certainly results in the absence of immune responses, whereas pMHC binding does not invariably trigger immune responses. As a result, predicting non-epitope regions is viewed as a more straightforward task.

This trend is also observed when predicting T-cell epitopes on complete Fel d 1 sequences. We also performed both prediction methods on the full sequences of Chain 1 and 2. Broadly speaking, the IEDB prediction method tends to over-predict peptides, and no direct correlation could be found between the epitope score and T-cell epitopes. Conversely, the pMHC binding prediction results from ProPred largely align with the known Fel d 1 T-cell epitopes (**Fig. 3**).

Discussion

In this study, we benchmarked various computational methods for prediction in Fel d 1 epitopes, highlighting the feasibility and potential of these tools in identifying allergenic epitopes. Even though our analysis focused on Fel d 1, these computational tools can be applied to other allergens. A comprehensive understanding of the epitope landscape of various allergens could contribute to the development of effective allergen-specific therapies and diagnostic tools.

While computational methods show promise, they have limitations in their ability to accurately identify epitopes. Our results demonstrate that these methods are in general more effective at predicting non-epitopes than epitopes. Specifically, most B-cell epitope methods we tested were only marginally better than random selection. This outcome may be due to these methods not being specifically designed to identify IgE epitopes. ElliPro, a structure-based B-cell epitope prediction tool, outperformed most of the other methods, suggesting the benefit of considering the 3D structure of allergenic proteins when predicting IgE epitopes.

In terms of T-cell epitope prediction, ProPred successfully identified all known T-cell epitopes for their associated alleles in our analysis. However, the IEDB prediction method failed to predict two known T-cell epitopes. The high rate of over-prediction observed with the IEDB method further emphasizes the challenges associated with predicting T-cell epitopes.

Drawing from our results, we anticipate that the development of advanced artificial intelligence

techniques, coupled with 3D structure prediction methods, could notably enhance the precision and specificity of epitope prediction tools. Such techniques are currently underdeveloped for allergen epitope identification. Additionally, expansive and comprehensive databases of experimentally validated epitopes would facilitate these tools in refining their algorithms, leading to improved epitope prediction.

In conclusion, our study illuminates both the potential and limitations of current computational methods in allergenic epitope prediction. Despite certain shortcomings, these computational tools still remain useful for generating preliminary insights into potential allergenic regions within proteins.

References

1. Chan SK, Leung DY. Dog and cat allergies: current state of diagnostic approaches and challenges. Allergy, Asthma & Immunology Research. 2018;10(2):97-105.
2. van Ree R, van Leeuwen WA, Bulder I, Bond J, Aalberse RC. Purified natural and recombinant Fel d 1 and cat albumin in in vitro diagnostics for cat allergy. Journal of allergy and clinical immunology. 1999;104(6):1223-30.
3. Arbes Jr SJ, Cohn RD, Yin M, Muilenberg ML, Friedman W, Zeldin DC. Dog allergen (Can f 1) and cat allergen (Fel d 1) in US homes: results from the National Survey of Lead and Allergens in Housing. Journal of allergy and clinical immunology. 2004;114(1):111-7.
4. Pomés A. Relevant B cell epitopes in allergic disease. International archives of allergy and immunology. 2010;152(1):1-11.
5. Cui Y. Immunoglobulin E-binding epitopes of mite allergens: From characterization to immunotherapy. Clinical reviews in allergy & immunology. 2014;47:344-53.
6. Vailes LD, Li Y, Bao Y, DeGroot H, Aalberse RC, Chapman MD. Fine specificity of B-cell epitopes on Felis domesticus allergen I (Fel d I): effect of reduction and alkylation or deglycosylation on Fel d I structure and antibody binding. Journal of allergy and clinical immunology. 1994;93(1):22-33.
7. Bond JF, Brauer AW, Segal DB, Nault AK, Rogers BL, Kuo M-C. Native and recombinant Fel dI as probes into the relationship of allergen structure to human IgE immunoreactivity. Molecular immunology. 1993;30(16):1529-41.
8. van Milligen FJ, van't Hof W, van den Berg M, Aalberse RC. IgE epitopes on the cat (Felis domesticus) major allergen Fel d I: a study with overlapping synthetic peptides. Journal of allergy and clinical immunology. 1994;93(1):34-43.
9. De Groot AS. Immunomics: discovering new targets for vaccines and therapeutics. Drug discovery today. 2006;11(5-6):203-9.
10. Sanchez-Trincado JL, Gomez-Perosanz M, Reche PA. Fundamentals and methods for T-and B-cell epitope prediction. Journal of immunology research. 2017;2017.
11. El-Manzalawy Y, Honavar V. Recent advances in B-cell epitope prediction methods.

Immunome research. 2010;6:1-9.
12. Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. Proceedings of the National Academy of Sciences. 1981;78(6):3824-8.
13. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. Journal of molecular biology. 1982;157(1):105-32.
14. Karplus P, Schulz G. Prediction of chain flexibility in proteins: a tool for the selection of peptide antigens. Naturwissenschaften. 1985;72(4):212-3.
15. Emini EA, Hughes JV, Perlow D, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of virology. 1985;55(3):836-9.
16. Chen J, Liu H, Yang J, Chou K-C. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino acids. 2007;33:423-8.
17. Yao B, Zhang L, Liang S, Zhang C. SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. 2012.
18. Singh H, Ansari HR, Raghava GP. Improved method for linear B-cell epitope prediction using antigen's primary sequence. PloS one. 2013;8(5):e62216.
19. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic acids research. 2017;45(W1):W24-W9.
20. Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome research. 2006;2(1):1-7.
21. Söllner J, Mayer B. Machine learning approaches for prediction of linear B-cell epitopes on proteins. Journal of Molecular Recognition: An Interdisciplinary Journal. 2006;19(3):200-8.
22. Yang H, Zhou Y, Cheng B, editors. Prediction of Linear B-cell Epitopes using Manifold Adaptive Experimental Design and Random Forest Algorithm. 2021 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS); 2021: IEEE.
23. Ras-Carmona A, Pelaez-Prestel HF, Lafuente EM, Reche PA. BCEPS: A web server to predict linear B cell epitopes with enhanced immunogenicity and cross-reactivity. Cells. 2021;10(10):2744.
24. Fukunaga K, Narendra PM. A branch and bound algorithm for computing k-nearest neighbors. IEEE transactions on computers. 1975;100(7):750-3.
25. Cai D, He X. Manifold adaptive experimental design for text categorization. IEEE Transactions on Knowledge and Data Engineering. 2011;24(4):707-19.
26. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS computational biology. 2012;8(12):e1002829.
27. Ansari HR, Raghava GP. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. Immunome research. 2010;6(1):1-9.
28. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins: Structure, Function, and Bioinformatics. 2006;65(1):40-8.
29. Jespersen MC, Mahajan S, Peters B, Nielsen M, Marcatili P. Antibody specific B-cell epitope predictions: leveraging information from antibody-antigen protein complexes. Frontiers in immunology. 2019;10:298.
30. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. Nucleic acids research. 2019;47(D1):D339-D43.
31. Greenbaum JA, Andersen PH, Blythe M, Bui HH, Cachau RE, Crowe J, et al. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. Journal of Molecular Recognition: An Interdisciplinary Journal. 2007;20(2):75-82.
32. Kaiser L, Velickovic TC, Badia-Martinez D, Adedoyin J, Thunberg S, Hallen D, et al. Structural characterization of the tetrameric form of the major cat allergen Fel d 1. Journal of

molecular biology. 2007;370(4):714-27.
33. Bonnet B, Messaoudi K, Jacomet F, Michaud E, Fauquert J-L, Caillaud D, et al. An update on molecular cat allergens: Fel d 1 and what else? Chapter 1: Fel d 1, the major cat allergen. Allergy, Asthma & Clinical Immunology. 2018;14(1):1-9.
34. Seppälä U, Hägglund P, Wurtzen PA, Ipsen H, Thorsted P, Lenhard T, et al. Molecular characterization of major cat allergen Fel d 1: expression of heterodimer by use of a baculovirus expression system. Journal of Biological Chemistry. 2005;280(5):3208-16.
35. Kaiser L, Gronlund H, Sandalova T, Ljunggren H-G, van Hage-Hamsten M, Achour A, et al. The crystal structure of the major cat allergen Fel d 1, a member of the secretoglobin family. Journal of Biological Chemistry. 2003;278(39):37730-5.
36. Kwok WW, Roti M, DeLong JH, Tan V, Wambre E, James EA, et al. Direct ex vivo analysis of allergen-specific CD4+ T cells. Journal of Allergy and Clinical Immunology. 2010;125(6):1407-9. e1.
37. 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 bioinformatics. 2008;9:1-8.
38. Clifford JN, Høie MH, Deleuran S, Peters B, Nielsen M, Marcatili P. BepiPred-3.0: Improved B-cell epitope prediction using protein language models. Protein Science. 2022;31(12):e4497.
39. Garnier J. Prediction of the secondary structure of proteins from their amino-acid sequence. J Mol Biol. 1978;120:97-120.
40. Pellequer J-L, Westhof E, Van Regenmortel MH. Correlation between the location of antigenic sites and the prediction of turns in proteins. Immunology letters. 1993;36(1):83-99.
41. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS letters. 1990;276(1-2):172-4.
42. Parker J, Guo D, Hodges R. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. 1986;25(19):5425-32.
43. Singh H, Raghava G. ProPred: prediction of HLA-DR binding sites. Bioinformatics. 2001;17(12):1236-7.
44. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, et al. Peptide binding predictions for HLA DR, DP and DQ molecules. BMC bioinformatics. 2010;11:1-12.
45. Southwood S, Sidney J, Kondo A, del Guercio M-F, Appella E, Hoffman S, et al. Several common HLA-DR types share largely overlapping peptide binding repertoires. The Journal of Immunology. 1998;160(7):3363-73.
46. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenetics. 2011;63:325-35.
47. Duffort O, Carreira J, Lombardero M. Monoclonal antibodies against Fel d I and other clinically relevant cat allergens. Immunology letters. 1988;17(1):71-7.

Acknowledgments

This study was financially supported by Chonnam National University (Grant number: 2021-2216) and National Research Foundation of Korea grants funded by the Korean government (NRF-2020R1A5A2031185, NRF-2020M3A9G3080281, and 2021R1F1A1063769).

Figures and Tables

Table 1. T-cell epitopes of Fel d 1 allergen. As the default sequence length for the IEDB prediction method is 15, we extended the input sequence when its length was shorter than 15 residues. Neither of the prediction methods included HLA-DRB1*14:01 in their allele sets and thus no prediction was possible for the specific peptide (ENALSLLDKIYTS).

Chain	HLA type	Range	T-cell epitope	IEDB input sequence	ProPred	IEDB
1	DRB5*01:01	19-31	DEYVEQVAQY KAL	TP**DEYVEQVAQYK AL**PV	O	O
	DRB1*01:01	25-44	VAQYKALPVV LENARILKNC	**VAQYKALPVVLEN ARILKNC**	O	O
	DRB1*13:01	31-43	LPVVLENARIL KN	KA**LPVVLENARILK N**CV	O	X
	DRB1*14:01	55-67	ENALSLLDKIY TS	DK**ENALSLLDKIYT S**PL	NA	NA
	DRB1*11:01	58-67	LSLLDKIYTS	DKENA**LSLLDKIYT S**PLC	O	X
2	DRB1*04:01	22-31	ELLLDLSLTK	VANGN**ELLLDLSLT K**VNATE	O	O

Table 2. Prediction results of computational B-cell epitope prediction tools for Fel d 1 IgE epitopes. Overall, he majority of methods yield results that are only marginally better than, or indistinguishable from, random selection. See also **Fig. 2** for other details.

	Prediction of Epitopes			Prediction of Non-epitopes		
	Precision	Recall	F1	Precision	Recall	F1
ElliPro	0.34	0.48	0.40	0.79	0.68	0.73
BepiPred-2.0	0.32	0.64	0.43	0.81	0.53	0.64
BepiPred-3.0	0.28	0.95	0.43	0.90	0.15	0.26
Emini	0.27	0.33	0.30	0.75	0.68	0.71
Kolaskar	0.27	0.55	0.37	0.76	0.49	0.60
BepiPred-1.0	0.22	0.21	0.22	0.73	0.73	0.73
Parker	0.25	0.45	0.32	0.74	0.53	0.62
Karplus	0.19	0.31	0.24	0.69	0.55	0.61
Chou	0.22	0.38	0.28	0.70	0.52	0.60

Figure 1. The structures and epitopes of Fel d 1. Chain 1 is colored in red and Chain 2 is in skyblue. (A) The heterodimeric structure of Fel d 1 (PDB ID: 1PUO). The three disulfide bonds are highlighted in a small sphere representation. (B) The tetrameric structure of Fel d 1 (PDB ID: 2EJN). Two Fel d 1 heterodimers are non-covalently bonded between two Chain 2s. (C) Three IgE epitopes of Fel d 1. (D) The residues that are included in the six T-cell epitopes (see **Table 1**). Overall, there are three main T-cell epitope regions.

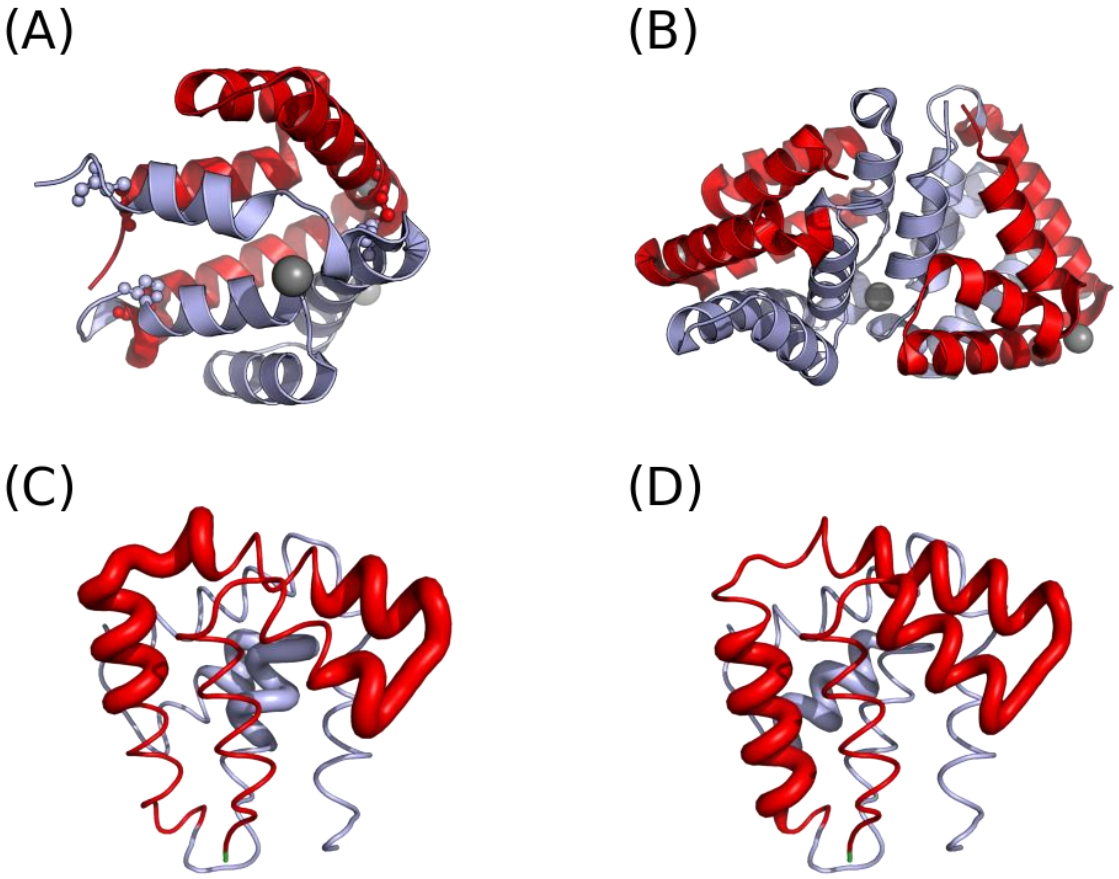


Figure 2. IgE epitope prediction using computational tools. Apart from BepiPred-2.0, 3.0, and ElliPro, all other methods are essentially no different from random selection. ElliPro, a structure-based prediction method, emerges as the best performer. See also **Table 2** for other details.

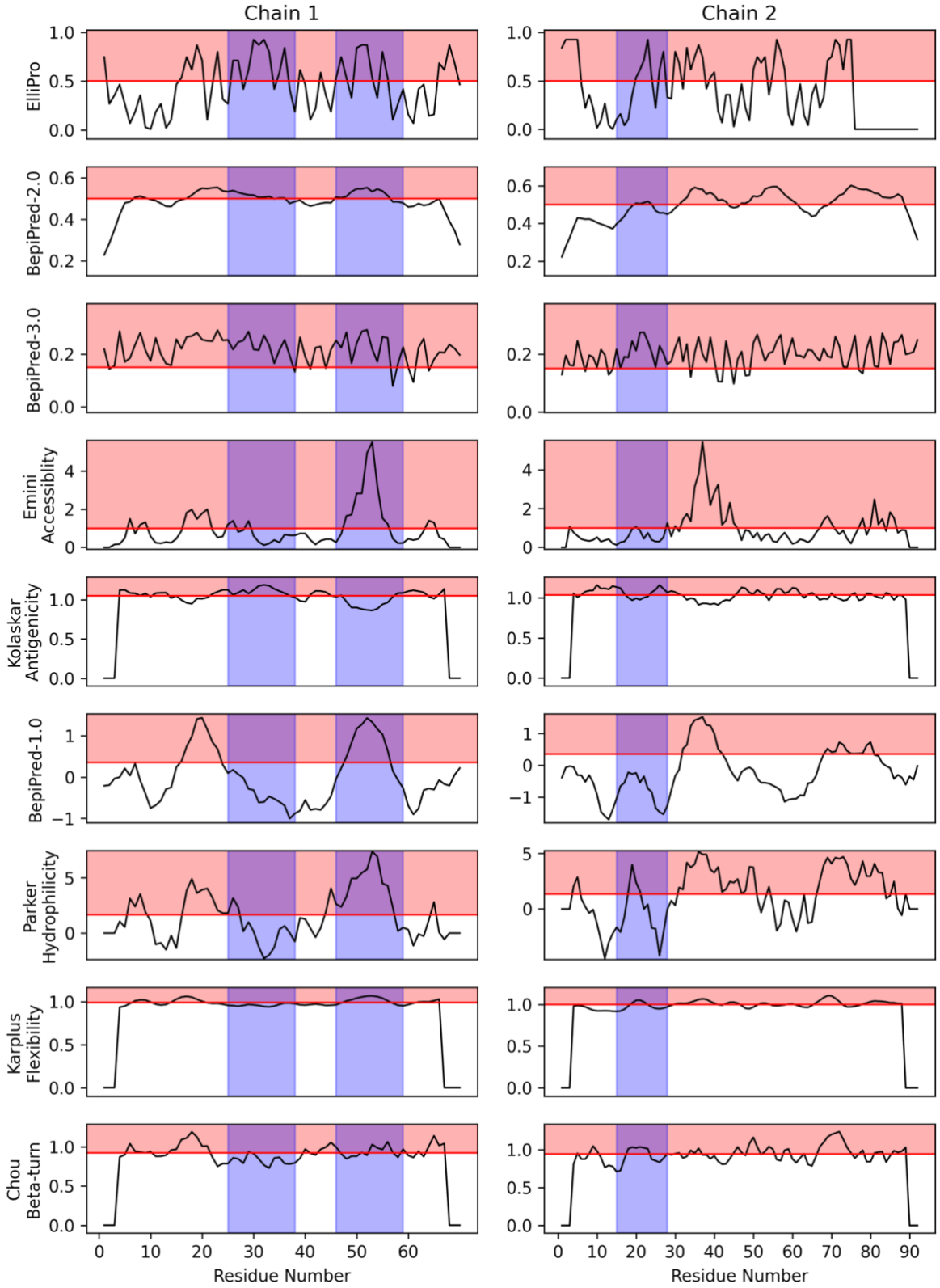
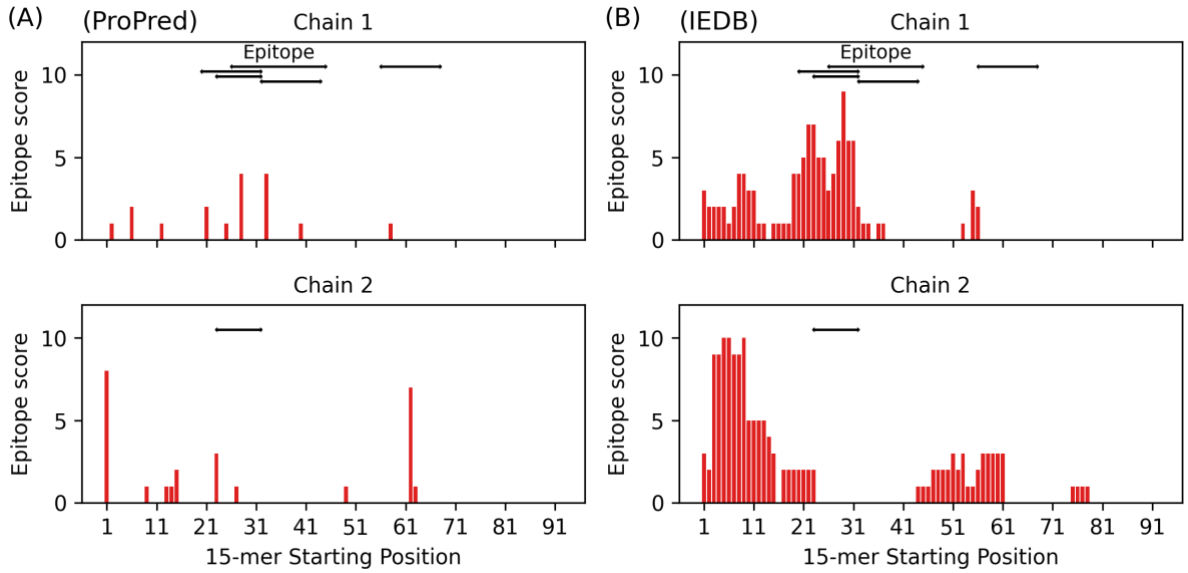
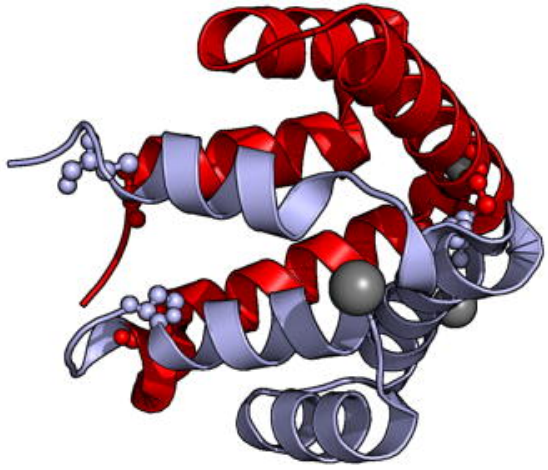


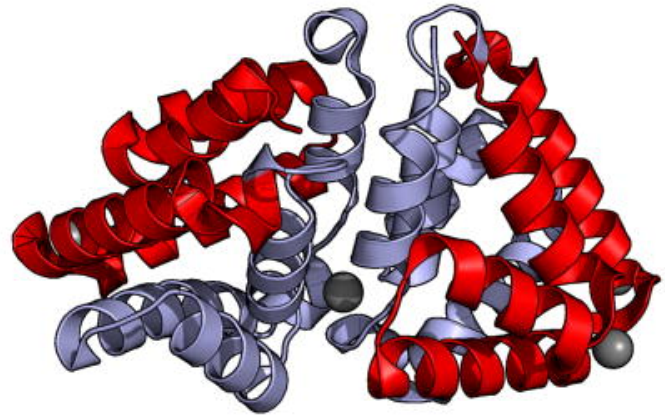
Figure 3. Fel d 1 T-cell epitope maps and prediction results. (A) The binders predicted by ProPred largely align with the actual T-cell epitopes. (B) The IEDB prediction method tends to overpredict potential T-cell epitopes.



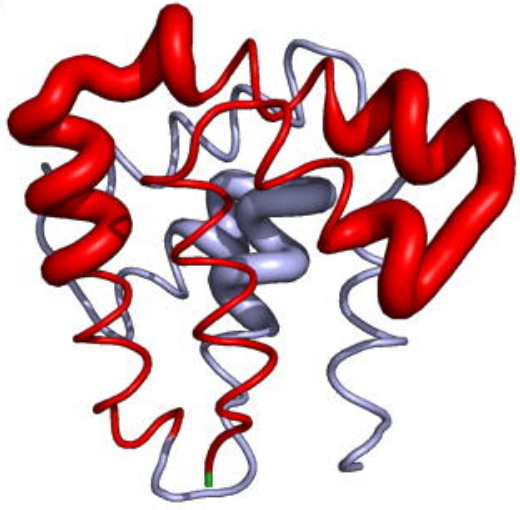
(A)



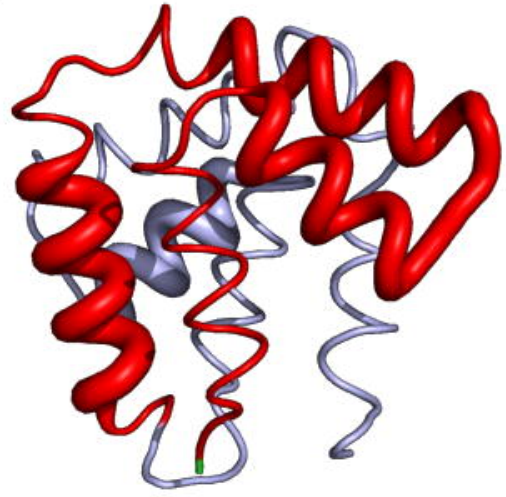
(B)

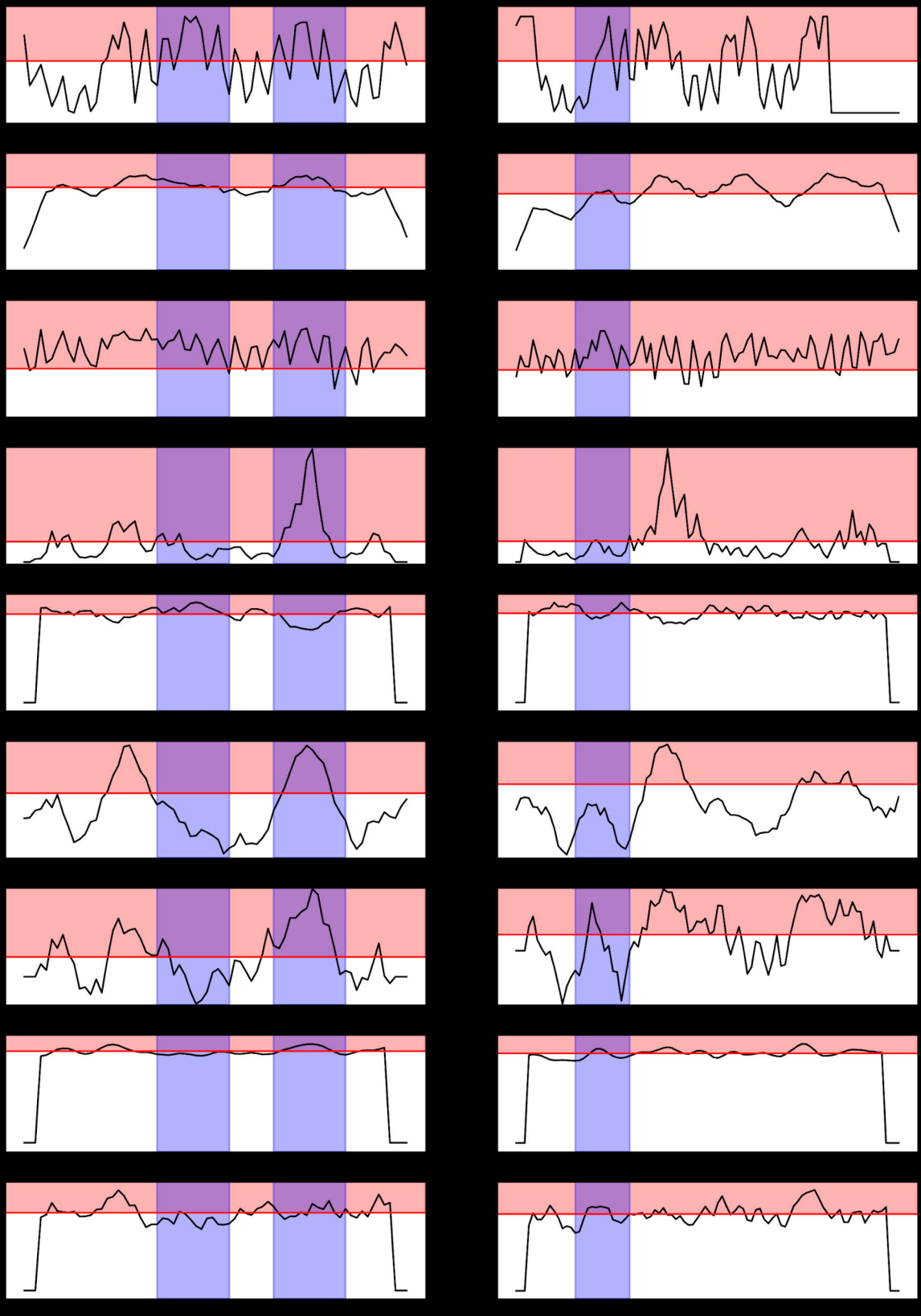


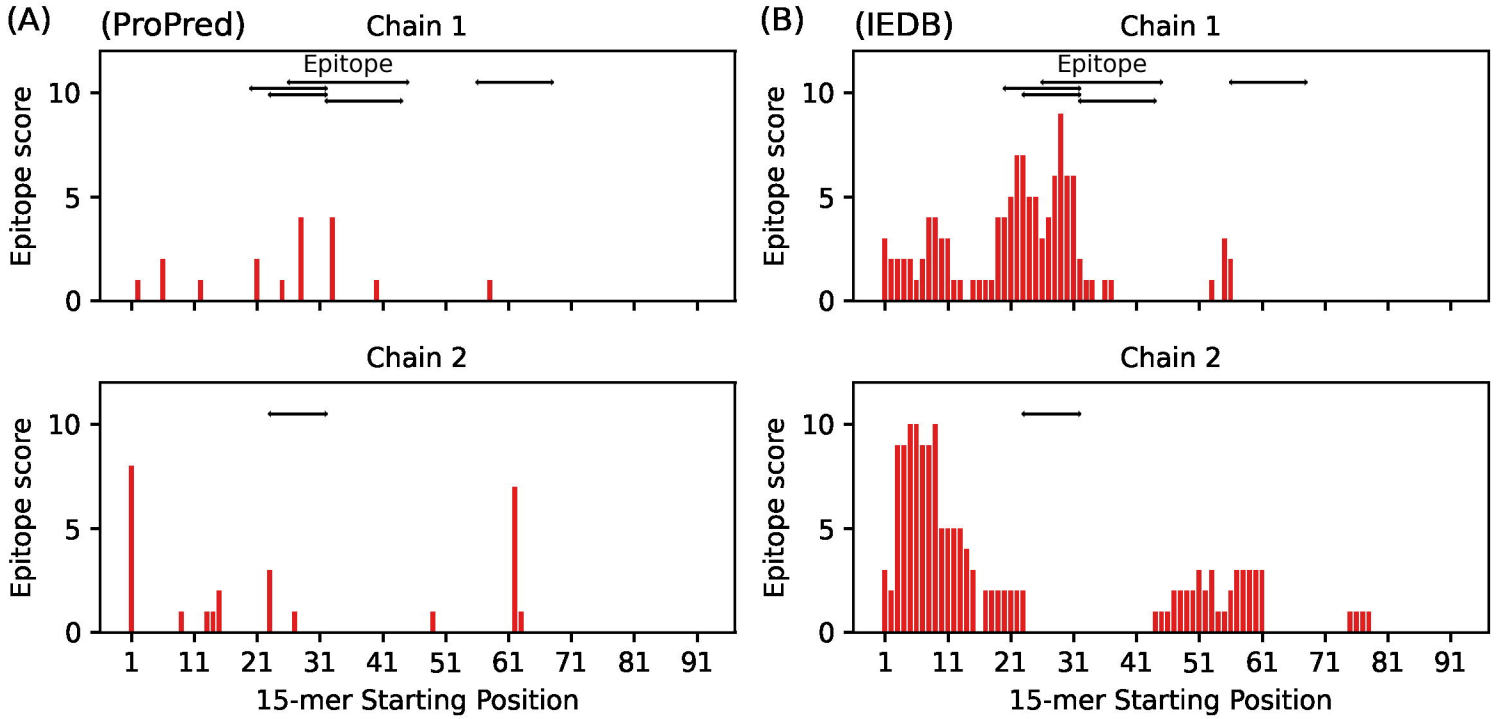
(C)



(D)






(A)
(ProPred)
Chain 1
Epitope
Epitope score
10
5
0
Chain 2
Epitope score
10
5
0
1
11
21
31
41
51
61
71
81
91
15-mer Starting Position
(B)
(IEDB)
Chain 1
Epitope
Epitope score
10
5
0
Chain 2
Epitope score
10
5
0
1
11
21
31
41
51
61
71
81
91
15-mer Starting Position