HHS Public Access

Author manuscript

Methods Mol Biol. Author manuscript; available in PMC 2021 May 10.

Published in final edited form as:

Methods Mol Biol. 2017 ; 1484: 255–264. doi:10.1007/978-1-4939-6406-2_17.

In Silico Prediction of Linear B-Cell Epitopes on Proteins

Yasser EL-Manzalawy, **Drena Dobbs**, **Vasant G. Honavar**

Abstract

Antibody-protein interactions play a critical role in the humoral immune response. B-cells secrete antibodies, which bind antigens (e.g., cell surface proteins of pathogens). The specific parts of antigens that are recognized by antibodies are called B-cell epitopes. These epitopes can be *linear*, corresponding to a contiguous amino acid sequence fragment of an antigen, or *conformational*, in which residues critical for recognition may not be contiguous in the primary sequence, but are in close proximity within the folded protein 3D structure.

Identification of B-cell epitopes in target antigens is one of the key steps in epitope-driven subunit vaccine design, immunodiagnostic tests, and antibody production. In silico bioinformatics techniques offer a promising and cost-effective approach for identifying potential B-cell epitopes in a target vaccine candidate. In this chapter, we show how to utilize online B-cell epitope prediction tools to identify linear B-cell epitopes from the primary amino acid sequence of proteins.

Keywords

Antibody-protein interaction; B-cell epitope prediction; Linear B-cell epitope prediction; Epitope mapping; Epitope prediction

1 Introduction

Antibodies, which are glycoproteins produced in membrane-bound or secreted form by B lymphocytes, mediate specific humoral immunity by engaging various effector mechanisms that serve to eliminate the bound antigens [1]. The characterization of antibody-protein interactions has been the focus of extensive research. This work has advanced our understanding of the adaptive immune system and contributed to important practical applications, such as identifying subunit vaccine targets [2, 3]. When an antibody binds to a protein, the resulting binding sites in the antibody and the protein are called the paratope and epitope, respectively. Among the several experimental methods for mapping B-cell epitopes and paratopes [2, 3], X-ray crystallography is perhaps the most preferred method because of its accuracy. Due to the high cost and technical challenges presented by experimental methods for mapping epitopes and para-topes, there is an urgent need for reliable in silico methods for identifying binding sites in antibody-protein complexes [4].

B-cell epitopes are classified as either linear or conformational. Linear epitopes are fragments of continuous amino acids in the protein sequence. Conformational epitopes

consist of amino acid residues that may be separated in the protein primary sequence, but are brought into physical proximity via protein folding. Although more than 90 % of epitopes are estimated to be conformational [5], most experimental studies and computational methods focus on mapping linear B-cell epitopes.

In this chapter, we discuss different computational methods for predicting linear and conformational B-cell epitopes and outline procedures for in silico identification of linear B-cell epitopes from amino acid sequence. Because the predictive performance of individual linear B-cell prediction methods is far from satisfactory, we propose a procedure that combines predictions from multiple predictors to obtain more reliable consensus predictions. Our approach also uses known or predicted 3D structures of target proteins to filter out false predictions. Due to the very limited availability of *sequence-based* conformational B-cell epitope prediction tools, consensus predictions are not currently feasible at present. However, with anticipated increase in the amount of experimental data, further advances in predicting conformational epitopes can be expected.

2 Materials

2.1 Data

In this protocol, the query is the primary sequence of a target protein (e.g., vaccine candidate). This vaccine candidate may be determined based on a literature survey (e.g., [6]) or using reverse vaccinology tools [7–9]. In some cases, the user may focus on protein fragments reported in literature or found to be conserved based on a multiple sequence alignment of the target protein sequences from multiple strains of the pathogen.

2.2 Linear B-Cell Epitope Prediction Tools

Early computational methods for mapping linear B-cell epitopes in an amino acid sequence assumed some correlation between a certain single physicochemical property of an amino acid (e.g., hydrophilicity, flexibility, or solvent accessibility propensity) and the likelihood that the amino acid would be part of a linear B-cell epitope [10–12]. BcePred [13] predicts linear B-cell epitopes using a combination of physicochemical properties as opposed to propensity measures based on a single amino acid property. BepiPred [14] combines the hydrophilicity scale proposed by Parker et al. [12] with a Hidden Markov Model (HMM) predictor. All these methods provide *residue-based* predictions, in that they assign a score to each residue in the query protein sequence; the higher the score assigned to a residue is, the more likely it belongs to a linear B-cell epitope (*see* Fig. 1 for an example).

Alternatively, several machine learning methods classify amino acid peptide chains of specific lengths as either epitopes or non-epitopes. BCPred [15] predicts linear B-cell epitopes of length 12, 14, 16, 18, 20, 22 amino acids using a Support Vector Machine (SVM) classifier and a string kernel. FBCPred [16] is a variant of BCPred for predicting linear B-cell epitopes of virtually any length. COBEpro [17] uses a two-stage procedure for predicting linear B-cell epitopes. In the first stage, an SVM classifier is used to assign scores to fragments of the query antigen. In the second stage, a prediction score is assigned to each residue in the query antigen based on the SVM scores for the peptide fragments. LBtope

[18] provides improved predictions of linear B-cell epitopes by training classifiers using experimentally validated *non*-epitopes, whereas all previous methods used randomly sampled fragments from UniProt as the non-epitope training data. Recently, we showed that further improvements in the reliability of linear B-cell epitope predictions can be obtained by using ensemble classifiers that combine multiple linear B-cell epitope predictors [19].

2.3 Conformational B-Cell Epitope Prediction Tools

The problem of conformational B-cell epitope prediction can be defined as follows: Given the primary or the tertiary structure of a query protein, what are the interfacial residues involved in the complex formed between the query protein and an antibody. This is essentially a subproblem of the more general problem of protein-protein interface prediction [20, 21], where the goal is to identify interfacial residues in a query protein that form a complex with any other protein (including antibodies). Unfortunately, protein-protein interface predictors trained on large data sets of protein-protein interfaces are not sufficiently reliable for predicting antibody-protein interfaces [22].

Partly due to the small number of solved antibody-protein structures, relatively few methods for predicting conformational B-cell epitopes have been proposed in the literature. The performance of the available methods remains far from satisfactory [4, 22]. Table 1 summarizes current B-cell epitope prediction methods that are available in the form of freely accessible web servers or downloadable software packages. In this table, we have categorized B-cell epitope prediction methods as *sequence-based* or *structure-based*, according to whether the method accepts the primary sequence vs. the 3D structural coordinates of the query protein as input. We have also categorized the methods as *residue-based* or *patch-based. Residue-based* methods return a prediction score for each residue in the query protein. *Patch-based* methods decompose the surface of the query protein into patches and return a single prediction score for each patch. Each patch could be interpreted as an epitope of an antibody-protein complex.

The vast majority of available tools for predicting conformational B-cell epitopes are *structure-based* in that they require the solved/predicted unbound structure of the target protein as input to the predictor. Hence, their applicability is limited by the availability of an experimentally determined 3D structure (from the PDB [23]) or a homology model for the query protein (*see* Note 1). To address this limitation, BEST [24] and CBTOPE [25] have been proposed for predicting conformational B-cell epitopes using amino acid derived information.

All of the methods described in Table 1 are antibody-independent B-cell epitope prediction methods [26], in the sense that they do not take advantage of information about the binding antibody in predicting the antibody binding site on the antigen. Recently, some antibody-specific B-cell epitope prediction methods have been proposed (*see* Note 2). Antibody-specific B-cell epitope prediction methods are motivated in part by: (1) the success of

[1]In the absence of solved 3D structure for a query protein, computational tools like I-TASSER [40] could be used to predict the 3D structure of that protein. I-TASSER is a template-based method for protein structure and function prediction. The pipeline consists of four major steps: template identification, structure reassembly, atomic model construction, and final model selection.

partner-specific protein-protein interface predictors [27, 28] and allele-specific major histocompatibility complex (MHC) binding site predictors [29, 30]; and (2) the observation that virtually any surface accessible region of an antigen can become the target of *some* antibody and elicit an immune response [26, 31] and hence it is much more useful to focus on the binding site for a specific antibody.

3 Methods

In this section, we focus on *sequence-based* tools for identifying linear B-cell epitopes.

3.1 Predicting Linear B-Cell Epitopes

Given the amino acid sequence of a protein of interest, apply the following procedure to obtain a list of predicted linear B-cell epitopes within the query sequence:

1. Go to submission page of BCPREDS server (*see* Fig. 2) accessible at http://ailab.ist.psu.edu/bcpred/predict.html.
2. Paste the amino acid sequence of the target protein.
3. Select the prediction method. The server currently supports three methods: BCPred [15], AAP [32], and FBCPred [16]. The user is encouraged to try all three methods (*see* **step 9**).
4. Select the length of the epitope. BCPred and APP methods can handle queries for a set of prespecified lengths (12, 14, 16, 18, 20, 22). FBCPred predicts linear B-cell epitopes of any length specified by the user. Some tips and guidelines for deciding on epitope length are provided in Note 3.
5. Select the specificity of the classifier (*see* Note 4).
6. Uncheck “report only non-overlapping epitopes” if you want the server to report all predicted epitopes with probability greater than the cut-off corresponding to the select classifier specificity in **step 6**. Otherwise, highly ranked non-overlapping epitopes will be also reported (*see* Note 5).
7. Click “Submit query” to obtain predicted epitopes in the query sequence.
8. Repeat **steps 1**–**8** for other supported prediction methods. Discard epitopes predicted by only a single method. The intuition behind this is that consensus

[2]Antibody-specific B-cell epitope prediction methods take into account the binding *antibody* sequence or structure in order to predict conformational B-cell epitopes in a query antigen sequence of known structure. EpiPred [34] is a fully *structure-based* method that requires the structures of an antigen and its putative binding antibody. Bepar [35] and ABepar [36] are fully *sequence-based* methods that take the sequences of the interacting antigen and antibody as input. PEASE server [37] predicts conformational B-cell epitopes in an antigen of known structure, given the sequence of the binding antibody.

[3]Deciding on optimal epitope length is not trivial. In fact existing tools cannot reliably predict optimal linear B-cell epitopes because most of the experimentally validated linear B-cell epitopes used to train these predictors are not optimal in length. However, it makes sense to use lengths between 12 and 16 amino acids because the lengths of known epitopes are within that range [15].

[4]There is always a trade-off between specificity and sensitivity. Higher specificity means lower false positive rate at the expense of missing some true positives (i.e., epitopes). We recommend using low specificity cut-offs and combining predictions from several tools to eliminate false positive predictions.

[5]A query protein sequence of L amino acids has L-k + 1 potential linear B-cell epitopes of length equal k. BCPREDS predictors assign a score to every candidate epitope and report epitopes with scores higher than the cut-off corresponding to user-specified specificity. To eliminate highly overlapping predicted epitopes and identify antigenic regions, the user might configure the tool to show non-overlapping epitopes.

predictions are usually more reliable than predictions obtained from a single prediction method.

9. Figure 3 shows the output of BCPREDS, in which non-overlapping epitopes predicted by the three prediction methods are combined and consensus predictions are identified (**bold** residues in the sequence).

10. Users are also encouraged to consider predictions by other servers (e.g., COBEPro [17]) by following essentially the same procedure described here to submit queries.

11. Evaluating the results: If possible, the user should filter out likely "false positives," i.e., predicted epitopes that do not lie on the surface of the protein by mapping the predicted epitopes onto a solved or predicted 3D structure of the query protein (*see* Note 6). In addition, the user might use the Immune Epitopes Database Analysis Resource (IEDB-AR) [33] to generate propensity scale profiles for the query protein (*see* Note 7). Although these profiles cannot provide reliable predictions of linear B-cell epitopes (*see* Note 8), they could be useful in highlighting potential antigenic regions of interest to confirm predictions by BCPREDS.

Acknowledgments

This work was supported by NIH grant GM066387 to VGH and DD, by Edward Frymoyer Chair of Information Sciences and Technology at Pennsylvania State University to VGH, and by a Presidential Initiative for Interdisciplinary Research (PIIR) award from Iowa State University to DD.

References

1. Abbas AK, Lichtman AH, Pillai S (2014) Cellular and molecular immunology: with student consult online access Elsevier Health Sciences, Philadelphia, PA
2. Abbott WM, Damschroder MM, Lowe DC (2014) Current approaches to fine mapping of antigen–antibody interactions. Immunology 142(4):526–535 [PubMed: 24635566]
3. Reineke U, Schutkowski M (2009) Epitope mapping protocols, vol 524, Methods in molecular biology Humana Press, New York
4. EL-Manzalawy Y, Honavar V (2010) Recent advances in B-cell epitope prediction methods. Immunome Res Suppl 2:S2
5. Walter G (1986) Production and use of anti-bodies against synthetic peptides. J Immunol Methods 88(2):149–161 [PubMed: 2420900]
6. Wu X, Li X, Zhang Q, Wulin S, Bai X, Zhang T, Wang Y, Liu M, Zhang Y (2015) Identification of a conserved B-cell epitope on duck hepatitis A type 1 virus VP1 protein. PLoS One 10(2):e0118041 [PubMed: 25706372]

[6]Interactive molecular viewers like JMol [38] and PyMol [39] take PDB coordinate files as input and allow user to visualize protein 3D structures and highlight particular amino acid residues and support scripts and plugins for other tasks (e.g., determine interface residues or finding and highlighting surface residues).

[7]The Immune Epitopes Database Analysis Resource (IEDB-AR) B-cell tool available at http://tools.iedb.org/bcell/ generates propensity scale profiles for submitted amino acid sequences using BepiPred [14] and five other propensity scales. Figure 1 shows example profiles generated for Ebola Virus GP protein (UniProt ID Q05320) using surface BepiPred and three propensity scales (accessibility [10], flexibility [11], and antigenicity [31]).

[8]Blythe and Flower [37] have conducted a comprehensive assessment of about 500 amino acid physicochemical propensity scales in predicting linear B-cell epitopes (using a data set of 50 proteins) and showed that the performance of the best method is only slightly better than random guessing. This result was the main motivation of the machine learning-based methods for predicting linear B-cell epitopes.

7. Palumbo E, Fiaschi L, Brunelli B, Marchi S, Savino S, Pizza M (2012) Antigen identification starting from the genome: a “Reverse Vaccinology” approach applied to MenB. In: Neisseria meningitidis: advanced methods and protocols. Methods in molecular biology, vol 799. Springer, pp 361–403
8. Donati C, Rappuoli R (2013) Reverse vaccinology in the 21st century: improvements over the original design. Ann N Y Acad Sci 1285(1):115–132 [PubMed: 23527566]
9. Xiang Z, He Y (2013) Genome-wide prediction of vaccine targets for human herpes simplex viruses using Vaxign reverse vaccinology. BMC Bioinformatics 14(Suppl 4):S2
10. Emini EA, Hughes JV, Perlow D, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 55(3):836–839 [PubMed: 2991600]
11. Karplus P, Schulz G (1985) Prediction of chain flexibility in proteins. Naturwissenschaften 72(4):212–213
12. Parker J, Guo D, Hodges R (1986) 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 25(19):5425–5432 [PubMed: 2430611]
13. Saha S, Raghava G (2004) BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In: Artificial immune systems. Lecture notes in computer science, vol 3239. Springer, pp 197–204
14. Larsen J, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. Immunome Res 2(2):1–7 [PubMed: 16426456]
15. EL-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting linear B-cell epitopes using string kernels. J Mol Recognit 21(4):243–255 [PubMed: 18496882]
16. EL-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting flexible length linear B-cell epitopes. In: Computational systems bioinformatics NIH Public Access, pp 121–132
17. Sweredoski MJ, Baldi P (2009) COBEpro: a novel system for predicting continuous B-cell epitopes. Protein Eng Design Select 22(3):113–120
18. Singh H, Ansari HR, Raghava GP (2013) Improved method for linear B-cell epitope prediction using Antigen’s primary sequence. PLoS One 8(5):e62216 [PubMed: 23667458]
19. EL-Manzalawy Y, Honavar V (2014) Building classifier ensembles for B-cell epitope prediction. In: Immunoinformatics. Methods in molecular biology, vol 1184. Springer, pp 285–294 [PubMed: 25048130]
20. Esmaielbeiki R, Krawczyk K, Knapp B, Nebel J-C, Deane CM (2015) Progress and challenges in predicting protein interfaces. Brief Bioinformatics bbv027
21. Xue LC, Dobbs D, Bonvin A, Honavar V (2015) Protein-protein interface predictions by data-driven methods: a review. FEBS Lett 589(23):3516–3526 [PubMed: 26460190]
22. Yao B, Zheng D, Liang S, Zhang C (2013) Conformational B-cell epitope prediction on antigen protein structures: a review of current algorithms and comparison with common binding site prediction methods. PLoS One 8(4):e62249 [PubMed: 23620816]
23. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28(1):235–242 [PubMed: 10592235]
24. Gao J, Faraggi E, Zhou Y, Ruan J, Kurgan L (2012) BEST: improved prediction of B-cell epitopes from antigen sequences. PLoS One 7(6):e40104 [PubMed: 22761950]
25. Ansari HR, Raghava G (2010) Identification of conformational B-cell epitopes in an antigen from its primary sequence. Immunome Res 6(6):1–9 [PubMed: 20167082]
26. Sela-Culang I, Ofran Y, Peters B (2015) Antibody specific epitope prediction—emergence of a new paradigm. Curr Opin Virol 11:98–102 [PubMed: 25837466]
27. Xue LC, Dobbs D, Honavar V (2011) HomPPI: a class of sequence homology based protein-protein interface prediction methods. BMC Bioinformatics 12(1):244 [PubMed: 21682895]
28. Minhas A, ul Amir F, Geiss BJ, Ben-Hur A (2014) PAIRpred: partner-specific prediction of interacting residues from sequence and structure. Proteins 82(7):1142–1155 [PubMed: 24243399]
29. El-Manzalawy Y, Dobbs D, Honavar V (2011) Predicting MHC-II binding affinity using multiple instance regression. Comput Biol Bioinformatics IEEE/ACM Trans 8(4):1067–1079

30. Trolle T, Metushi IG, Greenbaum JA, Kim Y, Sidney J, Lund O, Sette A, Peters B, Nielsen M (2015) Automated benchmarking of peptide-MHC class I binding predictions. Bioinformatics btv123

31. Rubinstein ND, Mayrose I, Halperin D, Yekutieli D, Gershoni JM, Pupko T (2008) Computational characterization of B-cell epitopes. Mol Immunol 45(12):3477–3489 [PubMed: 18023478]

32. Chen J, Liu H, Yang J, Chou K-C (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 33(3):423–428 [PubMed: 17252308]

33. Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, Bui H-H, Buus S, Frankild S, Greenbaum J (2008) Immune epitope database analysis resource (IEDB-AR). Nucleic Acids Res 36(suppl 2):W513–W518 [PubMed: 18515843]

34. Krawczyk K, Liu X, Baker T, Shi J, Deane CM (2014) Improving B-cell epitope prediction and its application to global antibody-antigen docking. Bioinformatics 30(16):2288–2294 [PubMed: 24753488]

35. Zhao L, Li J (2010) Mining for the antibody-antigen interacting associations that predict the B cell epitopes. BMC Struct Biol 10(Suppl 1):S6 [PubMed: 20487513]

36. Zhao L, Wong L, Li J (2011) Antibody-specified B-cell epitope prediction in line with the principle of context-awareness. Comput Biol Bioinformatics IEEE/ACM Trans 8(6):1483–1494

37. Sela-Culang I, Benhnia MR-E-I, Matho MH, Kaever T, Maybeno M, Schlossman A, Nimrod G, Li S, Xiang Y, Zajonc D (2014) Using a combined computational-experimental approach to predict antibody-specific B cell epitopes. Structure 22(4):646–657 [PubMed: 24631463]

38. Herraez A (2006) Biomolecules in the computer: Jmol to the rescue. Biochem Mol Biol Educ 34(4):255–261 [PubMed: 21638687]

39. DeLano WL (2002) Pymol: an open-source molecular graphics tool. CCP4 Newslett Protein Crystallogr 40:82–92

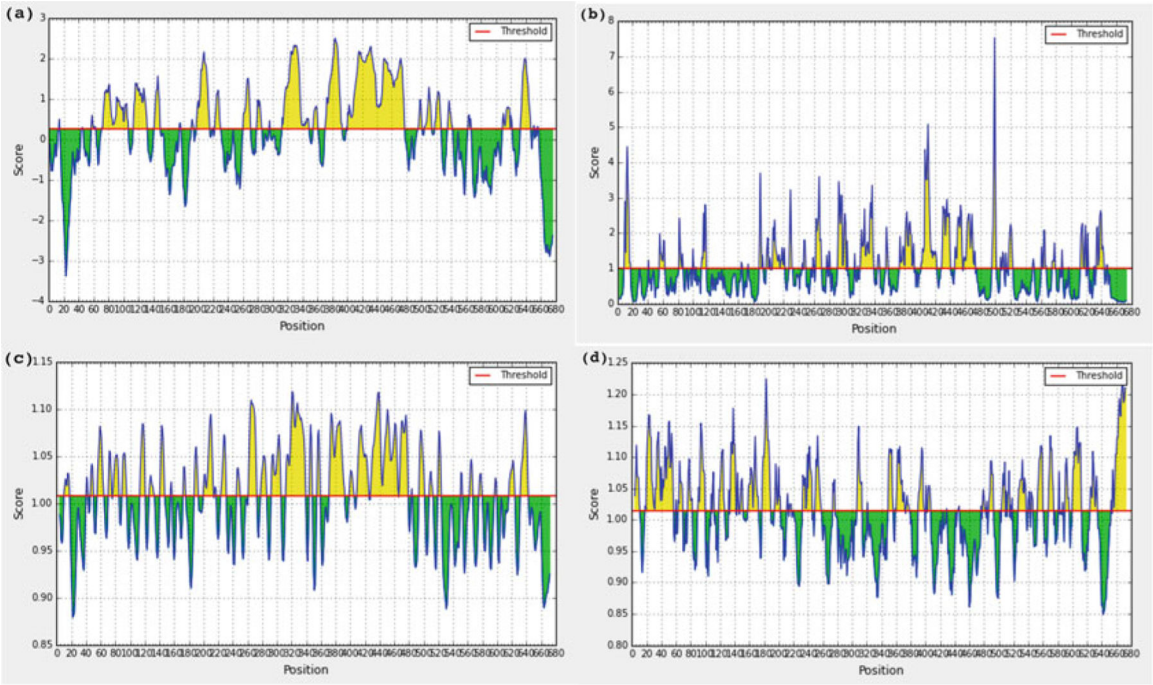

Fig. 1.

Propensity scale profiles for the Ebola virus GP protein (UniProt ID Q05320) generated using (**a**) BepiPred, (**b**) surface accessibility, (**c**) flexibility, (**d**) antigenicity. Regions with scores above the *red line* are more likely to contain linear B-cell epitopes

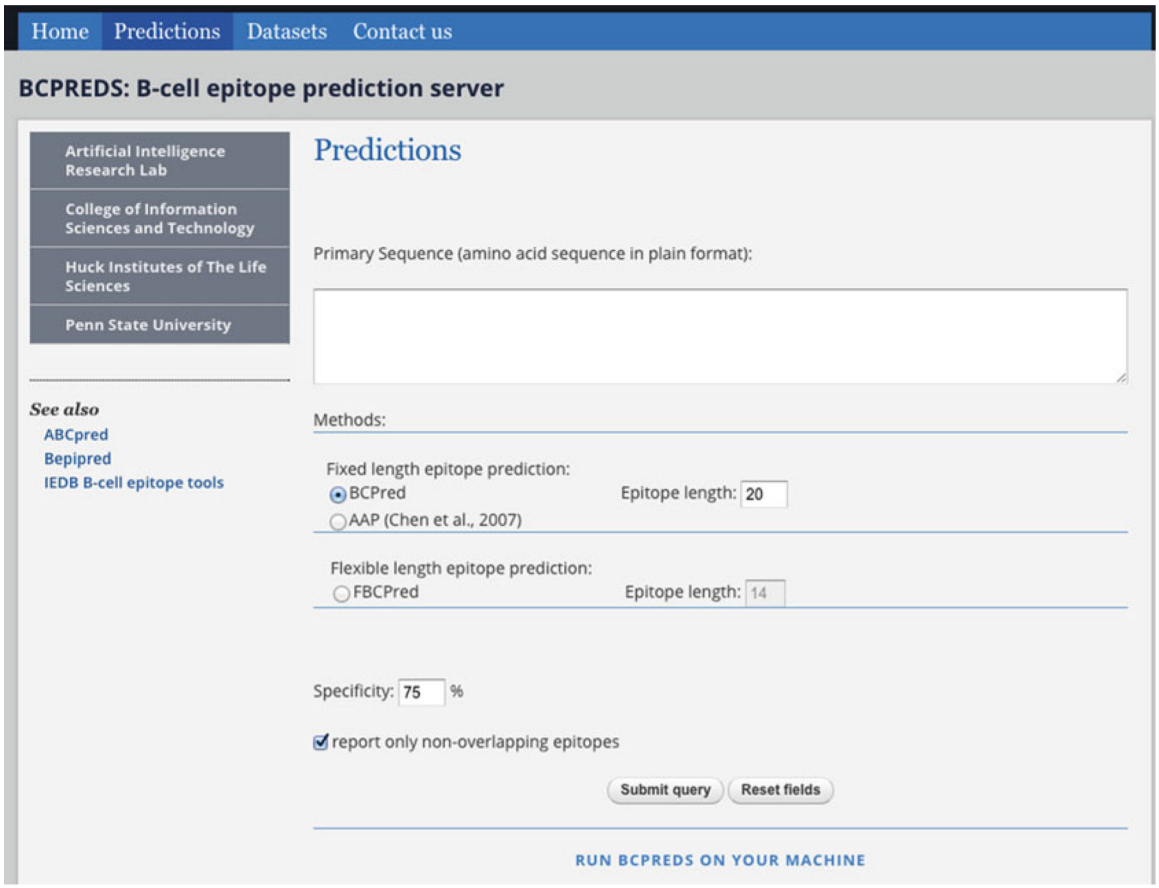

Fig. 2.

Submission page of BCPREDS web server available at: http://ailab.ist.psu.edu/bcpred/predict.html

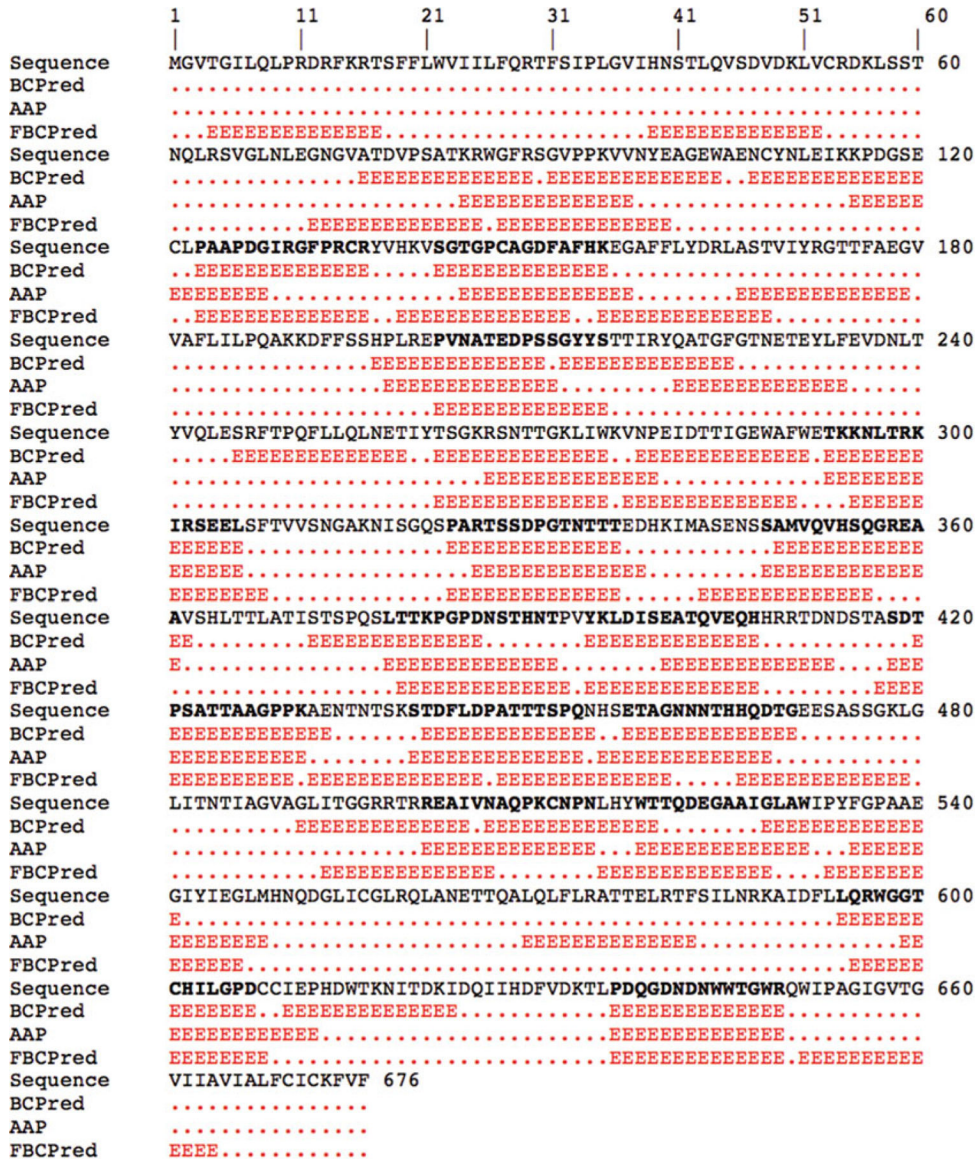

Fig. 3.

Linear B-cell epitopes predicted using three different linear B-cell epitope predictors currently supported by BCPREDS: BCPred, AAP, and FBCPred. Bold residues indicate epitope residues predicted by at least two methods

Table 1

Summary of antibody-protein binding site (conformational B-cell epitope) online prediction tools

Tool	URL of web server	Comments
CBTOPE	http://www.imtech.res.in/raghava/cbtope/	Sequence-based, residue-based
DiscoTope	http://www.cbs.dtu.dk/services/DiscoTope/	Structure-based, residue-based
ElliPro	http://tools.immuneepitope.org/ellipro/	Structure-based, residue-based
EPCES	http://sysbio.unl.edu/EPCES/	Structure-based, patch-based
Epitopia	http://epitopia.tau.ac.il/	Structure-based, residue-based
EPSVR	http://sysbio.unl.edu/EPSVR/	Structure-based, patch-based
PEPITO	http://pepito.proteomics.ics.uci.edu/	Structure-based, residue-based
SEPPA	http://lifecenter.sgst.cn/seppa2/	Structure-based, residue-based