Bioinformatics, 31(8), 2015, 1313–1315
doi: 10.1093/bioinformatics/btu790
Advance Access Publication Date: 27 November 2014
Applications Note

OXFORD

Structural bioinformatics

PEASE: predicting B-cell epitopes utilizing antibody sequence

Inbal Sela-Culang[1], Shaul Ashkenazi[1], Bjoern Peters[2] and Yanay Ofran[1,*]

[1]The Goodman Faculty of Life Sciences, Nanotechnology Building, Bar-Ilan University, Ramat-Gan 52900, Israel and [2]Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037, USA

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received on August 10, 2014; revised on November 4, 2014; accepted on November 22, 2014

Abstract

Summary: Antibody epitope mapping is a key step in understanding antibody–antigen recognition and is of particular interest for drug development, diagnostics and vaccine design. Most computational methods for epitope prediction are based on properties of the antigen sequence and/or structure, not taking into account the antibody for which the epitope is predicted. Here, we introduce PEASE, a web server predicting antibody-specific epitopes, utilizing the sequence of the antibody. The predictions are provided both at the residue level and as patches on the antigen structure. The tradeoff between recall and precision can be tuned by the user, by changing the default parameters. The results are provided as text and HTML files as well as a graph, and can be viewed on the antigen 3D structure.

Availability and implementation: PEASE is freely available on the web at www.ofranlab.org/PEASE.

Contact: yanay@ofranlab.org

1 Introduction

Antibody (Ab) epitopes, or B-cell epitopes, are used in many applications, such as diagnostics, therapy, vaccine design and biological research. Correct identification of Ab epitopes is crucial for all these applications and is also required for the understanding of Ab-Ag recognition and specificity (Sela-Culang *et al.*, 2013). Experimental methods for mapping Ab epitopes are expensive, laborious and time consuming, and many of them also fail to identify some of the epitopes (Xu *et al.*, 2010).

Developing an accurate computational method for epitope prediction has been proven to be a difficult task (Kringelum *et al.*, 2012; Yao *et al.*, 2013). Most methods attempt to differentiate between epitopic and non-epitopic surface residues of the Ag, based on properties associated with Ag sequence and/or structure (Yang and Yu, 2009). However, it appears that almost any residue on the Ag surface may become an epitope under some circumstances (Greenbaum *et al.*, 2007). Inspired by the success of T-cell epitope prediction methods, where predictions depend on the specific major histocompatibility complex molecule presenting the epitope (Desai and Kulkarni-Kale, 2014; Lundegaard *et al.*, 2012), we hypothesized that a B-cell epitope should be predicted for a certain Ab rather than for any Ab, and that the information from the Ab should be utilized to enable such Ab-specific predictions. We have shown that this approach can achieve substantially improved predictions when compared with predictions based on the Ag's features only (Sela-Culang *et al.*, 2014). Several strategies that take into account the Ab sequence or structure were proposed and introduced, in the last few years: Antibody Specific Epitope Prediction (ASEP) (Soga *et al.*, 2010) is an index, computed based on Ab-Ag residue preferences, used to narrow down candidate residues predicted by conventional methods. Bepar (Zhao and Li, 2010) and ABepar (Zhao *et al.*, 2011) are sequence-based methods predicting Ab-specific epitopes based on association patterns of Ab-Ag residues (Bepar) and on Ab-Ag preferences of individual residues and residue pairs (ABepar).

EpiPred (Krawczyk *et al.*, 2014) is a fully structure-based method combining geometric matching with Ab-Ag residue preferences.

Here, we present Predicting Epitopes using Antibody SEquence (PEASE), a web-server that predicts epitopes, provided an Ab sequence and its cognate Ag structure. For a detailed description of the PEASE algorithm, see Sela-Culang *et al.* (2014). Briefly, for a given pair of Ab sequence and Ag structure, all combinations of one residue from the Ab Complementarity Determining Regions (CDRs) and one residue from the Ag surface are evaluated by a machine-learning model trained on properties from 120 Ab-Ag complexes. Each such pair of residues is assigned a 'pair-score' representing the prediction that these Ag and Ab residues contact each other in reality. A 'residue-score' for an Ag surface residue is its highest 'pair-score', where a higher score means that the residue is predicted more strongly to be in that Ab's epitope. An optional additional step identifies surface patches on the Ag, which contain multiple residues with high 'residue-scores'. Each patch is assigned a 'patch-score' which is the average 'residue-score' of all residues in the patch that are above the 'cutoff-score' (see later). Higher 'patch-scores' represent stronger predictions. PEASE was successfully applied for blindly predicting the epitopes of 12 Abs of the Vaccinia virus, later confirmed experimentally by several methods, including X-ray crystallography (Sela-Culang *et al.*, 2014).

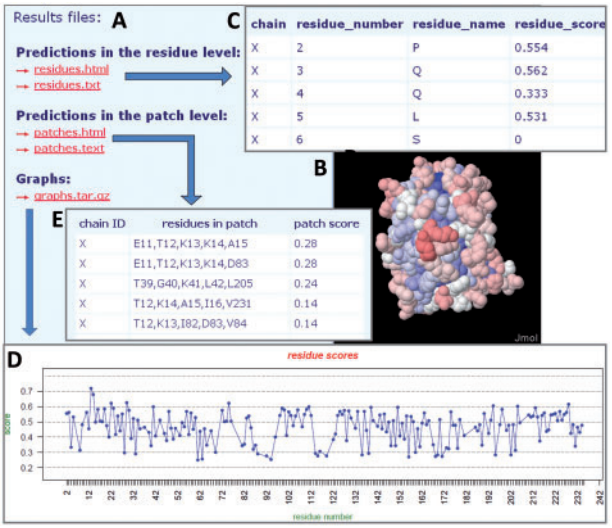

Fig. 1. An example of running PEASE to predict JE11 epitope on D8 surface. (**A**) the results page, (**B**) a jmol java applet of the Ag structure colored according to the 'residue-scores', (**C**) the 'residue-scores' HTML file, (**D**) a graph of the 'residue-scores' along the Ag sequence, (**E**) the patches HTML file

2 The PEASE server

The PEASE server contains two main pages: the 'software' page (in which the user uploads and specifies the input) and the results page. Additional tabs include an overview, a description of the input and output, a tutorial and FAQs. To better demonstrate PEASE, we describe it later, using as an example, the epitope prediction for D8 surface protein of the Vaccinia virus, and its JE11 Ab (which were not included in the training set; Sela-Culang *et al.*, 2014). The expected execution time for a submission is about 3 min for each Ag chain.

2.1 Inputs

On the 'software' page, the user should provide the Ag structure by either uploading the protein data bank (PDB) file or specifying the PDB code (4E9O in the D8 case). If the Ag structure is a computational model, the user should upload the Ag sequence file as well (to allow the PEASE algorithm to identify residues with no coordinates). Then, the user should specify for which Ag chain/s the prediction should be applied (chain X in the case of 4E9O structure). The predictions will be made separately for each of the selected chains. Finally, the Ab sequence should be uploaded or pasted in a text-box and should include both heavy and light chains. If the user provides a job name and an E-mail address, a message will be sent with a link to the results page, or to an error message, once the job is completed.

2.2 Parameters

The performance of prediction methods may be assessed by the recall and precision measures. Typically, there is a tradeoff between these two measures where one's improvement comes at the expense of the other. Depending on the objective of the epitope prediction, one may prefer a higher recall (i.e. reducing false negative) or higher precision (i.e. reducing false positive). The user may control this tradeoff by modifying the 'patch-size' and 'cutoff-score' parameters from their default values (5 and 0.6, respectively). Increasing the 'patch-size' will lead to a higher recall, but may result in lower precision. Increasing the 'cutoff-score' may increase precision, but may also reduce the number of predicted patches. We suggest starting with the default parameters, and then if necessary, modifying them slightly, one at a time.

2.3 Outputs

The PEASE results page (Fig. 1A) includes links to the results files (see later) and a jmol java applet showing the Ag 3D structure colored according to PEASE 'residue-scores' (Fig. 1B). Ag residues are colored from blue (low 'residue-score'—a low chance to be in the epitope) to red (high 'residue-score'—a high chance to be in the epitope). The user can rotate the colored Ag and consider the predictions in the context of the 3D structure. The PDB file of the Ag that contains the 'residue-scores' may be downloaded and viewed in any other protein 3D viewer. A pair of high scoring residues colored pink can be observed in the middle of Figure 1B. These residues (K13 and K14) were experimentally validated as being in the JE11 epitope (Sela-Culang *et al.*, 2014). The 'residue-scores' themselves are listed in text and HTML files (Fig. 1C shows the first N′ five residues). The 'residue-scores' may also be viewed along the Ag sequence as a graph (Fig. 1D). For the patches, a list of the patches and the corresponding 'patch-scores' is provided in text and HTML files. Figure 1E shows the first five patches, four of which were experimentally validated as being in the epitope. PEASE is modular, in the sense that one may use the file with the 'residue-scores' (Fig. 1C) as input for other methods that cluster residues based on a score and their spatial location, instead of considering the patches suggested by PEASE.

Acknowledgements

We thank Anat Burkovitz, Vered Kunik, Ariel Feiglin and Guy Nimrod for their help in designing the server.

Funding

This project has been funded in whole or in part with federal funds from the National Institutes of Allergy and Infectious Diseases [contract no: HHSN272200900048C].

Conflict of Interest: none declared.

References

Desai,D.V. and Kulkarni-Kale,U. (2014) T-cell epitope prediction methods: an overview. *Methods Mol. Biol.*, **1184**, 333–364.

Greenbaum,J.A. *et al.* (2007) Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J. Mol. Recognit.*, **20**, 75–82.

Krawczyk,K. *et al.* (2014) Improving B-cell epitope prediction and its application to global antibody-antigen docking. *Bioinformatics*, **30**, 2288–2294.

Kringelum,J.V. *et al.* (2012) Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput. Biol.*, **8**, e1002829.

Lundegaard,C. *et al.* (2012) Predictions versus high-throughput experiments in T-cell epitope discovery: competition or synergy? *Expert Rev. Vaccines*, **11**, 43–54.

Sela-Culang,I. *et al.* (2013) The structural basis of antibody-antigen recognition. *Front Immunol.*, **4**, 302.

Sela-Culang,I. *et al.* (2014) Using a combined computational-experimental approach to predict antibody-specific B cell epitopes. *Structure*, **22**, 646–657.

Soga,S. *et al.* (2010) Use of amino acid composition to predict epitope residues of individual antibodies. *Protein Eng. Des. Sel.*, **23**, 441–448.

Xu,X. *et al.* (2010) Evaluation of spatial epitope computational tools based on experimentally-confirmed dataset for protein antigens. *Chin. Sci. Bull.*, **55**, 2169–2174.

Yang,X.D. and Yu,X.L. (2009) An introduction to epitope prediction methods and software. *Rev. Med. Virol.*, **19**, 77–96.

Yao,B. *et al.* (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**, e62249.

Zhao,L. and Li,J.Y. (2010) Mining for the antibody-antigen interacting associations that predict the B cell epitopes. *BMC Struct. Biol.*, **10**, 56.

Zhao,L. *et al.* (2011) Antibody-specified B-cell epitope prediction in line with the principle of context-awareness. *IEEE-ACM Trans. Comput. Biol. Bioinform.*, **8**, 1483–1494.