Prediction and mapping of promiscuous MHC class II binders in an antigen sequence

Harpreet Singh
Indian Council of Medical Research Ansari Nagar, New Delhi - 110029, India
Gajendra Raghava
Institute of Microbial Technology, Sector 39A, Chandigarh, India

Method Article

Keywords: T-helper epitopes, promiscuous MHC binders, virtual matrices, prediction, subunit vaccine

Posted Date: November 7th, 2007

DOI: https://doi.org/10.1038/nprot.2007.502

Abstract

Introduction

There is tremendous change in strategy used for developing vaccine, over the years from whole pathogen to antigens and antigens to antigenic regions \(epitopes). In subunit vaccine epitope is used instead of a complete protein as vaccine candidate. Thus prediction of epitopes particularly T-help epitopes is one of the major challenges in subunit vaccine design. It is well established that binding of a peptide to an MHC Class II molecule is a prerequisite for activation of antigen specific T-helper cells. In past number of methods have been developed for predicting MHC class II binders but these methods allow to predict binder for one or two MHC alleles. Thus not suitable for wide population as individual have limited number of MHC alleles. In order to overcome this problem we developed a method for predicting promiscuous MHC binders, which binds two or more than two alleles. This server will be useful for designing subunit vaccine effective for wide population.

Reagents

Both formatted and non-formated sequences are accepted as input. For formatted sequences the server uses ReadSeq. software which can read most commonly used standard sequence formats including FASTA/PIR/EMBL/GENBANK etc. The user have to specify whether the sequence is in any format or non-formated as raw/plain text \(single letter coded amino acid only)

Equipment

User can access and use this web server from any computer \(Windows or Linux or Mac) with web browser and Internet connection

Procedure

To run prediction, follow these stepwise instructions. Step 1: Type the following URL address in your web browser "http://www.imtech.res.in/raghava/propred":http://www.imtech.res.in/raghava/propred Step 2: The user is required to fill the sequence submission form. A brief description of each of the field is as follows : Name of Antigen: This is an optional field. Paste your sequence below: Paste your antigen sequence in one of the standard format \(FASTA, EMBL, PIR etc.) or amino acid sequence only in single letter code. Or submit from file: The user can also upload the antigen sequence directly from a file. NOTE: Care should be taken that the server accepts input from either of two options, not both. Input sequence format: The user has to select the appropriate format according to the input sequence. Selection of parameters Threshold: The threshold is an important parameter which is defined as the 'percentage of best scoring natural peptides'. For example, a threshold of 1% would predict peptides in any given protein sequence which belong to the 1% best scoring natural peptides. The % threshold parameter allows the user to select for different stringency levels, in order to modulate the prediction results: a lower threshold

corresponds to a high stringency prediction, i.e. to a lower rate of false positives and to a higher rate of false negatives. In contrast, a higher threshold value \(low stringency) corresponds to a higher rate of false positives and a lower rate of false negatives. In short, from the same protein sequence input, a threshold setting of 1% will predict a lower number of peptide sequences and for a lower number of HLA-II alleles, compared to 2% or higher thresholds; however, this will also ensure a higher likelihood of positive downstream experimental results. Normally, at least for a first round of screening, threshold values higher than 3% are not desirable, since the rate of false positives can increase the size of the predicted repertoire to an amount unacceptable for later experimental testing. Display top scorer: Value in this field represent the number of top scorer in query antigen, to be displayed in tabular format. The peptide score of each nanomer in an antigen is calculated using quantitative matrices. The higher the score of any peptide frame the greater is the probability of it's binding to given MHC molecule. Default value is 5 % of the total number of nanomeric frames in query antigen. Allele: The user can select single/multiple allele form a list of 51 HLA-DR alleles. Multiple allele option is helpful in locating promiscuous binding regions Result Display Format: The server offers different result display formats to ease the identification of promiscuous binders. The server display the binding peptides in sequence for selected alleles. User can select desired format. Step 3: Finally click on "Submit" button.

Timing

1 minute for a protein

Troubleshooting

This server allows users to predict binders large number of alleles \(Upto 51). Server may take long time if you are predicting binders for large number of alleles.

Anticipated Results

HTML view I: Predicted binders are displayed as region underlined with " * " . This display is handy in locating overlapping binding regions in terms of their extend of overlap \(See Figure 1). HTML view II: Predicted binders are displayed as blue colored region, with P1 anchor or the starting residue of each predicted binding frame as red colored. This display is useful in locating promiscuous binding regions \(See Figure 2). Graphical View : The server allows to present the following parameters in graphical forScore distribution profile \(See Figure 3)

References

Singh,H. and Raghava,G.P.S.\(2001) ProPred: Prediction of HLA-DR binding sites. Bioinformatics,17\(12), 1236-37.

Acknowledgements

This work was supported by the Council of Scientific and Industrial Research and the Department of Biotechnology, Government of India.

Figures

Figure 1

Example output of PROPRED in HTML View I Click "here":http://protocols.nature.com/image/show/814 for a larger version of this image.

INPUT & PARAMETER INFORMATION

Antigen Name	antigen
Scanned on	Tue Nov 6 23:42:48 2007
Length of input sequence	140 amino acids
Number of nanomers from input sequence	132
Number of nanomers with obligatory P1 anchor residue	38
Threshold setting	5
Number of alleles in query	3

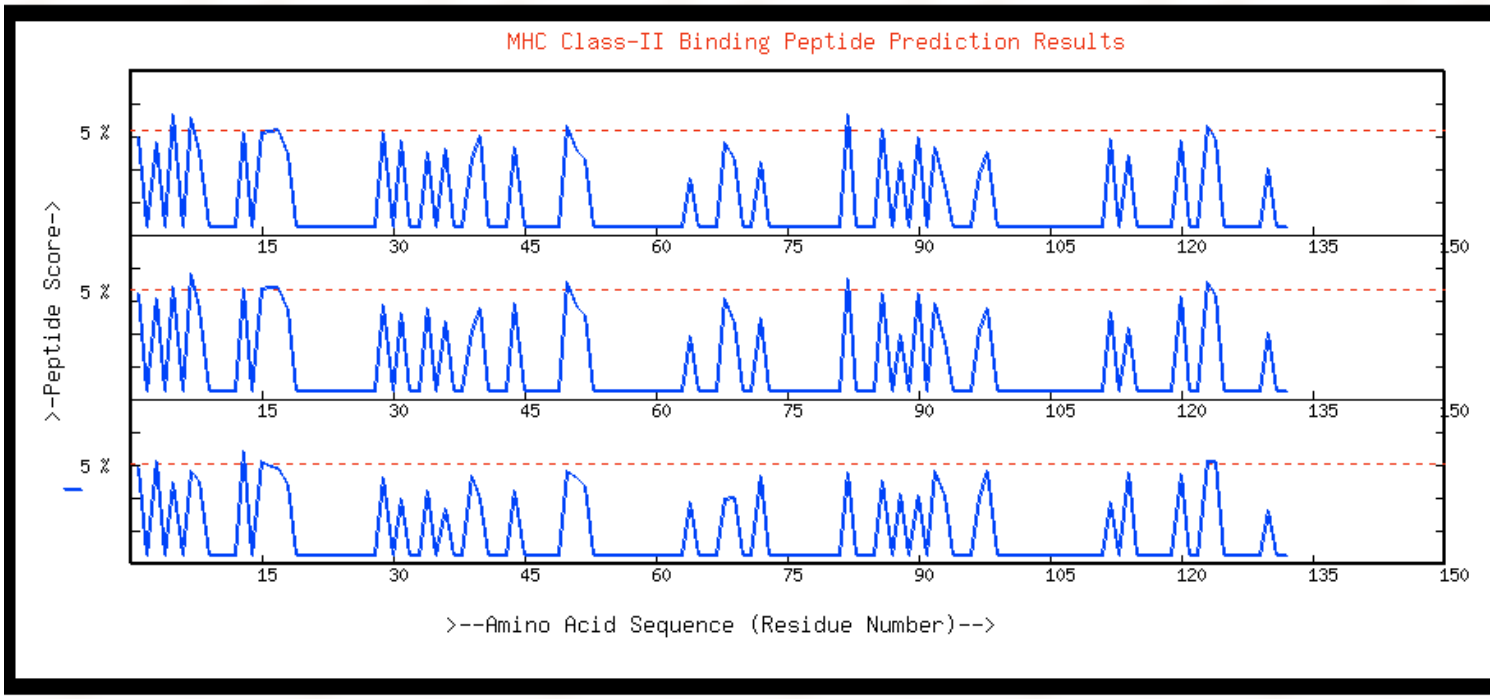
Subsequence Analysis from 1 till 12 GO For Predictions OR Reset

```
          ---------10--------20--------30--------40--------50--------60-
DRB1_0101: MKVKYALLSAGALQLLVVGCGSSHHETHYGYATLSYADYWAGELGQSRDVLLAGNAEADRA
DRB1_0102: MKVKYALLSAGALQLLVVGCGSSHHETHYGYATLSYADYWAGELGQSRDVLLAGNAEADRA
DRB1_0301: MKVKYALLSAGALQLLVVGCGSSHHETHYGYATLSYADYWAGELGQSRDVLLAGNAEADRA
```

GRAPHICAL DISPLAY | Sorted Binding Regions

Figure 2

Example output of PROPRED in HTML View II Click "here":http://protocols.nature.com/image/show/815 for a larger version of this image.



GRAPHICAL DISPLAY | Sorted Binding Regions

Figure 3

Example output of PROPRED in Graphical View Click "here":http://protocols.nature.com/image/show/816 for a larger version of this image.