IEDB-AR: immune epitope database-analysis resource in 2019

Dhanda, Sandeep Kumar; Mahajan, Swapnil; Paul, Sinu; Yan, Zhen; Kim, Haeuk; Jespersen, Martin C.; Jurtz, Vanessa; Andreatta, Massimo; Greenbaum, Jason A.; Marcatili, Paolo

Published in:
Nucleic acids research

Link to article, DOI:
10.1093/nar/gkz452

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
IEDB-AR: immune epitope database—analysis resource in 2019

Sandeep Kumar Dhanda1,†, Swapnil Mahajan1,†, Sinu Paul1,†, Zhen Yan1, Haeuk Kim1, Martin Closter Jespersen2, Vanessa Jurtz2, Massimo Andreatta2,3, Jason A. Greenbaum1, Paolo Marcatili2, Alessandro Sette1,4, Morten Nielsen2,3 and Bjoern Peters1,4,*

1Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037, USA, 2Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark, 3Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, Argentina and 4Department of Medicine, University of California, San Diego, CA 92122, USA

ABSTRACT

The Immune Epitope Database Analysis Resource (IEDB-AR, http://tools.iedb.org/) is a companion website to the IEDB that provides computational tools focused on the prediction and analysis of B and T cell epitopes. All of the tools are freely available through the public website and many are also available through a REST API and/or a downloadable command-line tool. A virtual machine image of the entire site is also freely available for non-commercial use and contains most of the tools on the public site. Here, we describe the tools and functionalities that are available in the IEDB-AR, focusing on the 10 new tools that have been added since the last report in the 2012 NAR webserver edition. In addition, many of the tools that were already hosted on the site in 2012 have received updates to newest versions, including NetMHC, NetMHCpan, BepiPred and DiscoTope. Overall, this IEDB-AR update provides a substantial set of updated and novel features for epitope prediction and analysis.

INTRODUCTION

The adaptive immune system in vertebrates can recognize a large repertoire of antigens from a broad spectrum of pathogens. B and T cell receptors are responsible for recognizing these diverse set of antigens and triggering immune responses. The specific regions recognized by these receptors are termed as epitopes. Thus, understanding the mechanism of immune receptor:epitope interactions is important in developing diagnostics, therapeutics, and vaccines against infectious and autoimmune diseases, cancers and allergies.

The Immune Epitope Database (IEDB) captures experiments that identify and characterize epitopes and epitope specific immune receptors along with various other details such as host organism, immune exposures, and induced immune responses (1). A companion site, IEDB-Analysis Resource (IEDB-AR), hosts various B and T cell epitope prediction tools based on algorithms trained and validated on the IEDB data along with epitope analysis tools. Since the last update, the number of monthly users visiting the IEDB-AR has more than tripled from under 1,500 in 2012 to over 4,500 in 2018 (Supplementary Figure S1). New epitope prediction and analysis tools are regularly added in the IEDB-AR with features to advance epitope-based therapeutics and vaccine development (2). For example, a tool to reduce undesired immunogenicity of therapeutic proteins was implemented recently (3). Here, we describe the newly implemented tools (Table 1), updates to the previously existing tools, and novel functionalities that have been added since the last report in the 2012 NAR webserver edition (4).

T CELL EPITOPE PREDICTION TOOLS

A total of 6 new tools were added in the category of T cell epitope prediction. These include TepiTool, a T cell peptide:MHC binding prediction tool with a new user-friendly interface, tools for prediction of naturally processed MHC class I and class II ligands, deimmunization of therapeutic proteins and prediction of T cell immunogenicity beyond MHC binding affinity. In addition to the newly added tools, many of the previously existing tools have been re-trained and updated as more data were made available. The latest versions of the prediction methods in T cell epitope prediction tools are listed in Table 2. While the latest versions are provided as the default methods, many of the tools allow the user to select previous versions where available. The newly added tools are described briefly in the following sections.

†To whom correspondence should be addressed. Tel: +1 858 752 6914; Fax:+1 858 752 6987; Email: bpeters@lji.org

© The Author(s) 2019. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
TepiTool (http://tools.iedb.org/tepitool) (5) is a new interface for IEDB T cell epitope predictions and is designed as a step-by-step wizard combining both MHC class I and class II prediction methods. The tool provides recommended default values at each step for the prediction and selection of an optimal set of peptides for a given application. TepiTool also offers additional functionalities that go beyond binding predictions. For example, conservancy analysis of peptides among the input sequences and different options for selecting the top-predicted peptides. Once the prediction task is finished, the user is provided with a concise set of top-predicted peptides and links to download the complete prediction results and conservancy estimates. The prediction results and links are also emailed to the user, if an email address is provided.

**Prediction of naturally processed ligands for MHC class I and class II**

MHC-NP (http://tools.iedb.org/mhcnp) (6) is a tool for predicting peptides that are naturally processed by the MHC class I pathway and bind to MHC class I molecules. The tool can predict MHC I ligands for six human and two mouse MHC alleles. Similarly, MHCII-NP (http://tools.iedb.org/mhcii-np) (7) is a tool for predicting naturally processed ligands for MHC class II. This new tool (http://tools.iedb.org/immunogenicity) is intended to classify peptides that bind to MHC class I (pMHC) into two categories: epitopes and non-epitopes (8). It is based on an analysis of amino acid composition of the peptide at non-anchor positions, where the side chains of amino acids are likely to be in contact with the TCR.

**Immunogenicity**

This new tool (http://tools.iedb.org/immunogenicity) is intended to classify peptides that bind to MHC class I (pMHC) into two categories: epitopes and non-epitopes (8). It is based on an analysis of amino acid composition of the peptide at non-anchor positions, where the side chains of amino acids are likely to be in contact with the TCR.

**CD4EpiScore**

CD4EpiScore (http://tools.iedb.org/cd4episcore) is a new tool for predicting the immunogenicity of CD4-restricted peptides in human populations that utilizes a neural network to identify patterns associated with immunogenicity.

### Table 1. New and updated tools in the IEDB-AR

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>Update type</th>
<th>Key features</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>T cell</td>
<td>TepiTool</td>
<td>New tool</td>
<td>Interactive and easy to use tool for immunologists</td>
<td>Prediction of T cell epitopes.</td>
</tr>
<tr>
<td></td>
<td>MHC-NP</td>
<td>New tool</td>
<td>Uses binding and ligand elution data to train the model.</td>
<td>Prediction of naturally processed ligands for MHC class I.</td>
</tr>
<tr>
<td></td>
<td>MHCII-NP</td>
<td>New tool</td>
<td>Uses motif information in the ligand elution dataset from IEDB</td>
<td>Prediction of naturally processed ligands for MHC class II.</td>
</tr>
<tr>
<td></td>
<td>Immunogenicity</td>
<td>New tool</td>
<td>Uses properties and position of amino acids to predict immunogenicity</td>
<td>Predicting immunogenicity for MHC-class I epitopes.</td>
</tr>
<tr>
<td></td>
<td>CD4EpiScore</td>
<td>New tool</td>
<td>Combines the prediction from immunogenicity and MHC binding algorithms</td>
<td>Predicting CD4 T cell reactivity in human population.</td>
</tr>
<tr>
<td></td>
<td>Deimmunization</td>
<td>New tool</td>
<td>Predicts non-immunogenic regions based on reduced binding to a set of reference MHC II alleles</td>
<td>Identification of immunogenic regions and suggested amino acid substitutions to reduce immunogenicity.</td>
</tr>
<tr>
<td>B cell / T cell</td>
<td>LYRA</td>
<td>New tool</td>
<td>Easy to use and fast antibody and TCR structure prediction.</td>
<td>Template-based 3D structure modeling of B- and T-cell receptors.</td>
</tr>
<tr>
<td>B cell</td>
<td>BepiPred2.0</td>
<td>New version</td>
<td>Training on conformational epitope dataset using random forest algorithm</td>
<td>Prediction of linear B-cell epitopes.</td>
</tr>
<tr>
<td>Analysis tools</td>
<td>DiscoTope2.0</td>
<td>New version</td>
<td>Novel spatial neighborhood and surface exposure definitions.</td>
<td>Prediction of discontinuous B-cell epitopes.</td>
</tr>
<tr>
<td></td>
<td>RATE</td>
<td>New tool</td>
<td>Infers HLA restriction by generating a matrix of subjects and given immune response</td>
<td>Inferring allele restriction for epitopes based on immune response data from HLA-typed subjects.</td>
</tr>
<tr>
<td></td>
<td>ImmunomeBrowser</td>
<td>New tool</td>
<td>User specified epitopes and source proteins</td>
<td>Aggregating and mapping the immune response from heterogeneous epitope data to source proteins.</td>
</tr>
<tr>
<td></td>
<td>Cluster2.0</td>
<td>Re-engineered</td>
<td>Multiple clustering methods and visualization.</td>
<td>Grouping and visualizing peptides similar in sequence.</td>
</tr>
</tbody>
</table>

### Table 2. Methods and versions available in the IEDB T cell epitope prediction tools

<table>
<thead>
<tr>
<th>MHC class</th>
<th>Prediction method</th>
<th>Versions available</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHC class I</td>
<td>IEDB consensus (Recommended*)</td>
<td>2.18 (default)</td>
<td>Moutafis et al. (22)</td>
</tr>
<tr>
<td></td>
<td>NetMHCpan</td>
<td>4.0 (default), 3.0, 2.8</td>
<td>Jurtz et al. (23)</td>
</tr>
<tr>
<td></td>
<td>NetMHC (also called ANN)</td>
<td>4.0 (default), 3.4</td>
<td>Andreatta and Nielsen (24)</td>
</tr>
<tr>
<td></td>
<td>SMMPMBC</td>
<td>1.0</td>
<td>Kim et al. (25)</td>
</tr>
<tr>
<td></td>
<td>SM</td>
<td>1.0</td>
<td>Peters and Sette (26)</td>
</tr>
<tr>
<td></td>
<td>Comlib_sijaden2008</td>
<td>1.0</td>
<td>Sidney et al. (27)</td>
</tr>
<tr>
<td></td>
<td>PickPocket</td>
<td>1.1</td>
<td>Zhang et al. (28)</td>
</tr>
<tr>
<td></td>
<td>NetMHCcons</td>
<td>1.1</td>
<td>Karosene et al. (29)</td>
</tr>
<tr>
<td></td>
<td>netMHCstabpan</td>
<td>1.0</td>
<td>Rasmussen et al. (30)</td>
</tr>
<tr>
<td>MHC II</td>
<td>IEDB consensus (Recommended*)</td>
<td>2.17</td>
<td>Wang et al. (31)</td>
</tr>
<tr>
<td></td>
<td>NetMHCpan</td>
<td>3.1</td>
<td>Andreatta et al. (32)</td>
</tr>
<tr>
<td></td>
<td>NN-align</td>
<td>2.2</td>
<td>Nielsen and Lund (33)</td>
</tr>
<tr>
<td></td>
<td>SMM-align</td>
<td>1.1</td>
<td>Nielsen et al. (34)</td>
</tr>
<tr>
<td></td>
<td>Combinatorial Library</td>
<td>1.0</td>
<td>Sidney et al. (27)</td>
</tr>
<tr>
<td></td>
<td>Sturniolo</td>
<td>1.0</td>
<td>Sturniolo et al. (35)</td>
</tr>
</tbody>
</table>

*Recommended methods can change based on regular benchmarking evaluations.

**Nucleic Acids Research, 2019, Vol. 47, Web Server issue W503**
It has been validated on a series of independent datasets reported in the literature for different ethnicities and diverse antigens using a variety of experimental approaches.

Deimmunization

The Deimmunization tool (http://tools.iedb.org/deimmunization) was added to IEDB-AR to address the issue of undesired immune reactivity to therapeutically important proteins. In a stepwise wizard, this tool makes use of the class II peptide:MHC binding prediction tools to predict potentially immunogenic regions in protein sequences and suggest amino acid substitutions to reduce their immunogenicity (3). As a proof-of-concept, the tool has been validated experimentally on recombinant factor VIIa (Vatreptacog alpha), which was discontinued from clinical trials due to immunogenicity issues (10).

B CELL EPITOPE PREDICTION TOOLS

The IEDB-AR hosts linear B cell epitope prediction tools, such as BepiPred (11), various amino acid physicochemical property based scales (http://tools.iedb.org/bcell/), and discontinuous B cell epitope prediction tools, such as DiscoTope (http://tools.iedb.org/discootope/) (12) and ElliPro (http://tools.iedb.org/ellipro/) (13). Since the last update of IEDB-AR, the recommended B cell epitope prediction methods, BepiPred (14) and DiscoTope (15), were updated to their 2.0 versions.

LYRA

A new tool named LYRA (Lymphocyte Receptor Automated Modelling) was added to model the 3D structures of B and T cell receptors (16). The LYRA tool (http://tools.iedb.org/lyra/) predicts the structure of B- and T-cell receptors from their amino acid sequence. Using homology modelling, it selects the best framework templates and, if necessary, models the complementary determining regions (CDRs) based on the predicted canonical structure (17) of each loop, which are then grafted onto the framework templates. The results page shows the aligned sequence and a visualization of the structure allowing for quick inspection of the CDRs in both sequence and structure.

ANALYSIS TOOLS

The analysis section of IEDB-AR contains tools that automate common tasks when working with sets of epitopes or epitope-candidates. Updates to this section include a revised version of epitope clustering along with a new tool to map epitopes to source proteins, and a tool to infer allele restrictions of epitopes from immune response data on HLA-typed subjects.

Cluster2.0

The epitope clustering tool (http://tools.iedb.org/cluster2) has been completely re-engineered to group peptidic epitopes based on their sequence similarity. In addition to providing three different clustering approaches, this new version also supports interactive graphical visualizations of the clusters to show connectivity among peptides (18).

ImmunomeBrowser

The ImmunomeBrowser tool in the IEDB website maps epitopes to their source antigen and provides a visualization of the observed immune responses across all tested regions of the protein. In a new customizable version of this tool (http://tools.iedb.org/immunomebrowser/), we have extended this application to perform a similar analysis for user-provided datasets of epitopes and source antigens (19).

Restrictor Analysis Tool for Epitopes (RATE)

RATE (http://tools.iedb.org/rate) (20) is an automated method that can computationally infer the HLA restrictions of epitopes, given large datasets of T cell responses in HLA typed subjects. The tool takes two inputs, the alleles expressed by the subject and the immune response of the peptides in the subjects. It then calculates the odds ratios for each allele being the restricting allele for a specific peptide and estimates significance using Fisher’s exact test. The tool was developed with a focus on class II alleles but can also be applied to class I alleles.

NEW FEATURES IN THE IEDB-AR

In addition to implementing new tools, IEDB-AR development since 2012 has also targeted improvements that address how users want to interact with the tools. Two of the most readily apparent and prevalent new features are the ability to submit prediction jobs for processing in the background and the architectural changes made at the hardware and software levels to improve stability and support parallelization.

Background batch job processing

With the release of version 2.17 of the IEDB-AR in June 2017, users were given the ability to provide an email address upon submission of a class I or class II peptide binding prediction job. This enhancement has allowed users to run larger prediction jobs (in terms of the number of input sequences and predicted alleles) than would be possible directly through the web interface. Upon completion of the job, results are sent as an email attachment to the user. Since the initial implementation, this feature has been added to TepiTool and the Deimmunization tool - both of which are computationally intensive and could timeout with a reasonable-sized request through the web interface. For tools that support batch processing, it is available through the web interface as well as the API.

Hardware & software architectural changes to improve stability and support parallelization

Several architectural changes have been made to the hardware and software in order to decouple the front-end from the back-end, improve stability, and support parallelization of several tools. At the hardware level, a separate job-processing cluster was created to run all CPU-intensive tasks, such as binding predictions. These machines are physically separated from the web server so that heavy processing has little effect on web site performance. To make use
of the redesigned server architecture, the backend software was completely reengineered in Python and Django with special attention to make use of a message queueing system (RabbitMQ) and task manager (Celery). With this integrated system in place, it has allowed parallelization of jobs for several of the resource tools with speedups as great as 15-fold over the single-threaded version. It has also enabled efficient use of resources and prioritization of jobs based upon their origins. As an example, a separate resource queue has been configured in collaboration with the Griffith lab to support the CPU-intensive requests of their pVAC-Seq pipeline (21) while keeping the IEDB systems responsive to requests from other users.

**AVAILABILITY OF THE TOOLS**

**IEDB application programming interface (IEDB-API)**

In addition to the main web interface, public-facing APIs are made available for several of the tools hosted at the IEDB-AR. Included among these tools are the MHC class I and class II binding and processing predictions, MHC-NP, and the B cell linear epitope predictor. To provide a consistent experience, each of the APIs adhere to a similar interface, with parameter names shared among them where possible. All of the APIs work via HTTP POST requests and return responses in plain text. The MHC class I and II binding prediction APIs are very heavily utilized, accounting for over 300,000 predictions each month and upward of 90% of the jobs run through the IEDB-AR.

**Software distribution packages**

While the goal of the public IEDB-AR server is to accommodate as many prediction jobs as reasonably possible, resources can be limiting for extremely large requests. This is one of the many reasons that the IEDB-AR team provides downloadable packages to run the predictions locally on the user hardware. Currently, 8 different standalone packages are available, covering the most widely used tools, and new packages are developed based upon user demand and available resources. Additionally, a complete virtual machine image of the IEDB-AR is made available to external entities through license agreements. These two modes of distribution cover a broad range of use cases, enabling users to run many of the IEDB-AR tools on their own hardware and in complete privacy.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**FUNDING**

National Institutes of Health [HHSN272201200010C, 75N93019C00001]. Funding for open access charge: National Institutes of Health [HHSN272201200010C, 75N93019C00001].

Conflict of interest statement. None declared.

**REFERENCES**


