

# Immune Epitope Database and Analysis Program

## IEDB User Documentation

La Jolla Institute for Allergy and Immunology  
9420 Athena Circle  
La Jolla, CA 92037

16 May 2017

**This page intentionally left blank**



2.4.4	Video Tutorials.....	45
2.5	More IEDB .....	46
2.5.1	Database Export .....	46
2.5.2	Meta-Analyses .....	48
2.5.3	Citing the IEDB .....	48
2.5.4	Release Notes .....	49
2.5.5	Links to External Sources .....	49
2.6	Learn More .....	49
2.6.1	Support .....	50
2.6.2	About the Data.....	51
2.6.3	About Us.....	51
3	Publications of the IEDB team by Year .....	52
3.1	2016.....	52
3.2	2015.....	53
3.3	2014.....	54
3.4	2013.....	54
3.5	2012.....	55
3.6	2011.....	56
3.7	2010.....	56
3.8	2009.....	57
3.9	2008.....	57
3.10	2007.....	58
3.11	2006.....	59
3.12	2005.....	59
3.13	2004.....	60
4	References.....	61

## Table of Figures

Figure 2.1 IEDB 3.6 Home page .....	2
Figure 2.2 Home page Simple Search .....	3
Figure 2.3 Query results for a substring search performed on the input sequence AELLVALENQHTIDL .....	4
Figure 2.4 Epitope Detail Search input screen .....	6
Figure 2.5 Epitope panel on the Specialized Search page expanded to show input fields.....	7
Figure 2.6 Epitope query fields in the Advanced Search for Linear peptide epitopes .....	8
Figure 2.7 Epitope query fields in the Advanced Search for Discontinuous peptide epitopes ....	8
Figure 2.8 Epitope query fields in the Advanced Search for Discontinuous peptide on multi- chain epitopes.....	9
Figure 2.9 Epitope query fields in the Advanced Search for Non-peptidic epitopes .....	9
Figure 2.10 MHC Assay Detailed Search web page.....	10
Figure 2.11 Identifier Search input screen.....	11
Figure 2.12 The Epitope tab of the search results page .....	12
Figure 2.13 An example of the epitope detail page .....	13
Figure 2.14 The Antigens tab of the search results page .....	14
Figure 2.15 The Results filtered by the Nucleoprotein of the Influenza A virus. This was achieved by clicking on the filter icon in the antigen column. The content of the other tabs contain data specific to this antigen.....	14
Figure 2.16 The Assays tab of the search results page. The page contains three additional tabs for T Cell, B Cell, and MHC Ligand Assays .....	15
Figure 2.17 The top three sections of a sample T cell assay detail page.....	16
Figure 2.18 The References tab of the search results page .....	17
Figure 2.19 An example of the Reference detail page.....	18
Figure 2.20 Sample MHC Allele Details page with Peptide MHC Binding Motif diagram for HLA- A*02:01. Note the link to the corresponding Amino Acid Binding Chart on the bottom left of the diagram.....	19
Figure 2.21 Amino acid binding matrix for MHC allele HLA-A*02:01 .....	20
Figure 2.22 The Immunome Browser is accessed by clicking on the icon shown on the Antigen tab on the results page of a query .....	21
Figure 2.23 A pop-up window appears each time the user clicks on the Antigen tab on the Results page. It highlights the Immunome Browser icon and briefly explains the Immunome Browser's functionality. ....	22
Figure 2.24 Immunome Browser graphical output displaying assay results mapped onto a reference proteome by residue position .....	23
Figure 2.25 Immunome Browser tabular output.....	23
Figure 2.26 Allele Finder. Alleles are displayed in a tree structure and users can search by allele name, source organism, and MHC class (I, II, or non-classical).....	25
Figure 2.27 Assay Finder on the T Cell Advanced Search page. Assays are displayed in a tree structure and users can find displays by their method/technique and by the characters with which the assay name starts .....	26
Figure 2.28 The disease finder is particularly helpful in specifying epitopes related to autoimmune diseases. In the example shown, diabetes is highlighted and selected. ....	27
Figure 2.29 The non-peptidic version of the Molecule Finder. This finder is accessed in the Epitope filter on the search results page. ....	28
Figure 2.30 An example of the protein branch of the Molecule Finder.....	29
Figure 2.31 Organism Finder showing a search for dengue with dengue highlighted in the tree .....	31

Figure 2.32 Geolocation finder. This example shows the tree structure by continent, region, and country. A search was made for Mexico within the tree using the Search By box on the left, which produced one result. This was then highlighted in the tree, as shown. Clicking on “Mexico” in the tree populated the Current Selection field in the upper left corner of the finder. 32

Figure 2.33 IEDB Database Export web page .....48

Figure 2.34 The Learn More page, accessed from the Welcome message in the upper left corner of the IEDB home page .....50

# 1 Introduction

The Immune Epitope Database and Analysis Resource (IEDB) is a public repository of immune epitope data and is sponsored by the National Institute for Allergy and Infectious Diseases (NIAID). The IEDB development started in December 2003 and it became available to the public in a beta test phase on 15 February 2006. The IEDB contains data related to antibody and T cell epitopes for humans, non-human primates, rodents, and other animal species. The IEDB also makes available a variety of analytical and epitope prediction tools and resources within its Analysis Resource.

## 2 Website Features

The IEDB website functionality can be divided into five categories— Home Page (Section 2.1), Specialized Searches (Section 2.2.2), Analysis Resource (Section 2.3), Help (Section 2.4), and More IEDB (Section 2.5), which largely correspond to the pull-down menus at the top of the home page. The subsections that follow describe the website features within these categories.

The reader will find it helpful to keep in mind how data are stored in the IEDB. Each item contained in the IEDB consists of a reference (article or submission) containing information about one or more epitopes and associated binding or response information. The same epitope can exist in multiple references.

### 2.1 Home Page

The IEDB Home Page ([www.iedb.org](http://www.iedb.org)) is the default screen displayed when users enter the IEDB system. Besides providing a general description of the IEDB project, the home page displays system level status and notification of scheduled updates or maintenance. As one can see in Figure 2.1, the page is divided into three columns that contain introductory information, Summary Metrics of the data, a basic data search capability, and links to the epitope prediction and analysis tools in the Analysis Resource. As users browse the IEDB system, they can return to the home page anytime by clicking *Home* on the far left of the main menu bar or by clicking on the IEDB logo in the upper left-hand corner of the page.

The top box in the left column contains a welcome statement that briefly describes the content of the database. It also has a “Learn More” link that brings the user to a web page containing more information about the support and user help features available, about the data itself, and about the IEDB program. This page is described in detail in Section 2.6. The Summary Metrics are displayed in the bottom box of this column. These numbers are intended to be a gauge of the volume of data available in the system. The Summary Metrics provide the number of epitopes (peptidic and non-peptidic), assays (T cell, B cell, and MHC ligand assays), epitope source organisms, restricting MHC alleles, and references. A middle box may appear when there are special messages to convey to users. The text will appear in red. The middle column contains the initial search interface that is described in detail in Section 2.2.1. This section was updated at the end of March 2017 to include several finders on the search interface, thus enabling more complete queries without the need for further filtering of the results. The rightmost column has links to the most frequently used tools in the Analysis Resource.

**IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE**

Home | Specialized Searches | Analysis Resource | Help | More IEDB

**Welcome**

The IEDB is a free resource, funded by a contract from the National Institute of Allergy and Infectious Diseases. It offers easy searching of experimental data characterizing antibody and T cell epitopes studied in humans, non-human primates, and other animal species. Epitopes involved in infectious disease, allergy, autoimmunity, and transplant are included.

The IEDB also hosts tools to assist in [Learn More](#)

**IEDB analysis of the Zika virus available here (analysis updated on an ongoing basis).**

**Summary Metrics**

Peptidic Epitopes	262,081
Non-Peptidic Epitopes	2,476
T Cell Assays	310,749
B Cell Assays	389,491
MHC Ligand Assays	532,262
Epitope Source Organisms	3,557
Restricting MHC Alleles	734
References	18,151

**START YOUR SEARCH HERE**

**Epitope**

- Any Epitopes
- Linear Epitope
- Exact M:  Ex: S1INFEKL
- Discontinuous Epitopes
- Non-peptidic Epitopes

**Assay**

- Positive Assays Only
- T Cell Assays
- B Cell Assays
- MHC Ligand Assays
- Ex: neutralization

**Antigen**

Organism:  Ex: influenza, peanut

Antigen Name:  Ex: core, capsid, myosin

**Host**

- Any Host
- Humans
- Mice
- Non-human Primates
- Ex: dog, camel

**MHC Restriction**

- Any MHC Restriction
- MHC Class I
- MHC Class II
- MHC Nonclassical
- Ex: HLA-A\*02:01

**Disease**

- Any Disease
- Infectious Disease
- Allergic Disease
- Autoimmune Disease
- Ex: asthma, diabetes

**Epitope Analysis Resource**

**T Cell Epitope Prediction**

Scan an antigen sequence for amino acid patterns indicative of:

- MHC I Binding
- MHC II Binding
- MHC I Processing (Proteasome, TAP)
- MHC I Immunogenicity

**B Cell Epitope Prediction**

Predict linear B cell epitopes using:

- Antigen Sequence Properties

Predict discontinuous B cell epitopes using antigen structure via:

- Solvent-accessibility (DiscoTope)
- Protrusion (EliiPro)

**Epitope Analysis Tools**

Analyze epitope sets of:

- Population Coverage
- Conservation Across Antigens
- Clusters with Similar Sequences

[Provide Feedback](#) | [Help Request](#) | [Solutions Center](#) | [Tool Licensing Information](#)

Supported by a contract from the [National Institute of Allergy and Infectious Diseases](#), a component of the National Institutes of Health in the Department of Health and Human Services.

Data Last Updated: March 24, 2017

Figure 2.1 IEDB 3.6 Home page

## 2.2 Query

There are several ways users can find information in the IEDB. A family of advanced query-by-example searches can be accessed on the Search pull-down menu and consist of “B Cell Search”, “T Cell Search”, “MHC Ligand Search”, “Epitope Search”, and “Identifier Search”. This group of queries are referred to as Specialized Searches. A simplified search that is especially designed for the needs of most immunologists appears on the home page. This document will refer to this particular query as the Simple Search or Home Page Search. All of these methods return results in a common format, as described in Section 2.2.3. These methods are elaborated in the subsections below.

### 2.2.1 Perform a Home Page Search

The home page search feature available in the center of the home page (Figure 2.2) is designed to simplify the search process for many commonly asked queries by immunologists. It is the most frequently used query mode. The approach it takes is analogous to many travel and shopping websites. IEDB users can start their search by specifying a number of fundamental and common search parameters on the home page. On the subsequent results page, users can then apply additional and more specific filters. The epitope structure choices include linear epitopes, discontinuous epitopes, or non-peptidic epitopes. For the linear epitope search, the user can prescribe the type of search desired - an exact match to the entered sequence, a substring search (described below), or a homologous peptides at sequence at identity levels of 70%, 80%, or 90% as determined by BLAST. The IEDB categorizes all assays as T cell, B cell, or MHC ligand. The MHC ligand includes binding and elution assays. Users can also select the MHC restriction for non-B cell searches. Finally, users can select the host species and the disease.

The results of the query are displayed in four tabs – Epitopes, Antigens, Assays, and References. The Epitopes tab lists the epitope description, such as the amino acid sequence for linear epitopes, its associated antigen and source organism, the number of references in which it appears, and the number of assays in which it is used. The Antigen tab lists the antigen, its source organism, and the number of related epitopes, assays, and references. The Assays tab contains three tabs, one each for T cell, B cell, and MHC ligand assays. All tabs list the IEDB-specific assay identifier, the reference in which it appears, and the related epitope. The other fields listed are those relevant for each assay type. For example, MHC restriction does not appear for B cell assays. The References tab lists the IEDB-specific reference identifier, the authors, title, abstract, and year of publication or submission. In all cases, the user can click on the IEDB-specific identifier to drill down to get additional information about the epitope, assay, or reference.

**START YOUR SEARCH HERE**

**Epitope**

Any Epitopes  
 Linear Epitope  
 Discontinuous Epitopes  
 Non-peptidic Epitopes

Exact M:  Ex: SIINFEKL

**Assay**

Positive Assays Only  
 T Cell Assays  
 B Cell Assays  
 MHC Ligand Assays

Ex: neutralization

**Antigen**

Organism  
 Ex: influenza, peanut

Antigen Name  
 Ex: core, capsid, myosin

**MHC Restriction**

Any MHC Restriction  
 MHC Class I  
 MHC Class II  
 MHC Nonclassical

Ex: HLA-A\*02:01

**Host**

Any Host  
 Humans  
 Mice  
 Non-human Primates

Ex: dog, camel

**Disease**

Any Disease  
 Infectious Disease  
 Allergic Disease  
 Autoimmune Disease

Ex: asthma, diabetes

**Figure 2.2 Home page Simple Search**

Once the initial results are generated, the user can filter the results using additional filtering criteria. Auto-complete fields and associated hierarchical tree views reveal the available parameter choices. In this way, users can narrow their query to specific non-peptidic epitopes, assay types, MHC alleles, hosts, or diseases. Users can also specify article type (journal article and/or submission), the author, title, PubMed ID, and year. All filters that are applied are displayed at the top of the page. Results on most pages can be downloaded as comma separated value (CSV) files that can be read using a text editor or spreadsheet application. Users also have the ability to map positive and negative results to reference proteomes as a means of visualizing immunogenic regions using the Immunome Browser(Section 2.2.3.2).

An autocomplete feature is available for many of the fields on the lefthand side of the results page that contain additional filter criteria. This includes the non-peptidic epitope field in the Epitope box, the organism and antigen name fields in the Antigen box, the T cell, B cell, and MHC ligand assay fields in the Assay box, the specific MHC restriction field in the MHC Restriction box, the specific host field in the Host box, the specific disease field in the Disease box, and the author, title, and data fields in the Reference box. For example, as one types “hep” in the organism field in the Antigen box, several choices start to appear in a list below the text field, including “hepatitis C virus”. The user can click on a selection from the list or can continue typing to further narrow down choices. Likewise, a user may type “human” or “homo” to select “homo sapiens” in the Organism field. The list of matches includes scientific names and synonyms. These actions enable users to quickly specify their search parameters without using the finders. Use of the autocomplete circumvents the need to use the finders.

If the user decides not to use the autocomplete feature for a search field, they will need to use the finders. The epitope source can be prescribed for the source organism and source antigen by using the organism finder (Section 2.2.4.5) and molecule finder (Section 2.2.4.4), respectively. The user can decide whether to include B cell responses, T cell responses, and/or MHC ligand results in the search (at least one must be checked). The host organism, the MHC restriction, and the MHC class can also be specified with the help of the organism finder and the allele finder (Section 2.2.4.1). The fields using finders will allow multiple selections as search criteria. In these cases the selections are treated as a set. Records will be considered a match if they include at least one of the selected values in the set. The search is executed by clicking the Search button and the query results can be viewed on the updated results page described in Section 2.2.3.

In addition to finding peptides in the database that contain the specified amino acid sequence, the substring search also finds epitopes within the input sequence itself. For example, when the sequence AELLVALENQHTIDL is submitted for a substring search, the query results, shown in Figure 2.3, yields five peptide sequences. Three of them (the second, third, and fifth) contain the input sequence. The two others (the first and fourth) are substrings of the input sequence that are also epitopes contained in the IEDB.

5 Records Found							
Details	Epitope	Antigen	Organism	# References	# Assays		
1055	AELLVALEN	Hemagglutinin	Influenza A virus	2	5		
1056	AELLVALENQHTIDL	Hemagglutinin	Influenza A virus	1	1		
50489	QDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTD	Hemagglutinin	Influenza A virus	1	1		
124806	VALENQHTI	Hemagglutinin	Influenza A virus	1	1		
130384	YNAELLVALENQHTIDL	Hemagglutinin	Influenza A virus	1	1		

5 Records Found

Go To Records Starting At 1200

Export Epitopes Results

**Figure 2.3 Query results for a substring search performed on the input sequence AELLVALENQHTIDL**

## 2.2.2 Specialized Searches

The Specialized queries are based on a standard Query by Example (QBE) approach, which is a method of forming queries where a user can enter conditions for each data field they want included in the query. The Specialized queries allow users to define example criteria for each field in the system. As there are over 300 fields, the Specialized queries are both powerful and comprehensive.

The fully expanded input screen for the Epitope Details Search is shown in Figure 2.4. The user can specify criteria in the left-hand panels for Epitope and Reference. These sections expand to the right to expose more detailed criteria for the search, including autocomplete fields and finders. An example of the expanded Epitope box is found in Figure 2.5. Users can specify search criteria for the epitope, such as epitope type (e.g. peptide from protein, carbohydrate), source molecule and organism, and various reference details. Further explanation of the search terms can be found in the Curation Manual. Using the bottom Reference section, the user can search on several reference criteria, such as author name, article title, the IEDB Reference ID, keywords in the reference abstract, year, author affiliation, and reference type. References are either published literature articles that have been curated by the IEDB curation staff or data submitted directly by researchers. This query type returns T cell, B cell, and MHC ligand data.

Auto-complete functionality is available for all the finder fields, as denoted by the greyed text that provide example inputs. The “?” provides a brief explanation and a “Learn More” link that connects to the appropriate help page in the IEDB Solutions Center. Finders are described in further detail in Section 2.2.4.

Figure 2.5 shows the fields when “Structure Type - Any” is selected in the Epitope filter. The fields for the other epitope type selections are shown in Figure 2.6 (linear epitopes), Figure 2.7 (Discontinuous epitopes), Figure 2.8 (Discontinuous peptides on Multichain), and Figure 2.9 (Non-peptidic epitopes).

IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE

Home Specialized Searches Analysis Resource Use the Legacy Site Help More IEDB

Epitope Detailed Search

Reset Search

Epitope ID

Structure Type - Any I

Organism

Antigen Name

Epitope Referenc...

Epitope Structure Defines

Evidence Code for Sourc...

Epitope Name

Reference Start Position

Reference End Position

Reference Region

Comments

Data Location in Reference

Epitope Related ...

Related Object

Type - Any Type

Organism

Antigen Name

Reference

Author

Title

Reference Details

Reference ID

Abstract

Affiliations

Date (Year)

Type - Any

Pending Filters Reference Type: 0

Epitopes (413612) Antigens (25693) Assays (988554) References (16889)

Go To Records Starting At 1200 Export Epitopes Results

413612 Records Found Page 1 of 16545 25 Per Page

Details	Epitope	Antigen	Organism	# References	# Assays
123885	cardiolipin			323	1239
44920	NLVPMVAIV	65 kDa phosphoprotein	Human herpesvirus 5 (Human cytomegalovirus)	223	562
113645	MEVGHYRSPFSRVVHLYRNGK	Myelin-oligodendrocyte glycoprotein	Mus musculus (mouse)	176	970
20354	GILGFVFTL	Matrix protein 1	Influenza A virus	165	382
58560	SIINFEKL	Gal d 2	Gallus gallus (chicken)	142	367
112741	2,4-dinitrophenyl group			140	573
4602	ASNENMETM	Nucleoprotein	Influenza A virus	133	379
24786	HSLGKWLGHDFDKF			115	781
112742	2,4,6-trinitrophenyl group			107	354
130649	alpha-Gal epitope			107	451
130694	1-O- (alpha-D-galactosyl) - N-hexacosanoylphytosphingosine			104	435
20788	GLCTLVAML	mRNA export factor ICP27 homolog	Human herpesvirus 4 (Epstein Barr virus)	103	227
6435	CINGVCWTV	Genome polyprotein	Hepatitis C virus	99	265
48237	PKYVKQNTLKLAT	Hemagglutinin	Influenza A virus	96	343
32208	KLVALGINAV	Genome polyprotein	Hepatitis C virus	86	225
61086	SSIEFARL	Envelope glycoprotein B	Human herpesvirus 1 (Herpes simplex virus type 1)	86	304
53112	RAHYNIVTF	Protein E7	Alphapapillomavirus 9	83	222
30001	KAVYNFATC	Pre-glycoprotein polyprotein GP complex	Lymphocytic choriomeningitis virus	82	266
7493	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA	Amyloid beta A4 protein	Homo sapiens (human)	81	301
6568	CLGGLLTMV	Latent membrane protein 2	Human herpesvirus 4 (Epstein Barr virus)	80	215
67436	TYQRTRALV	Nucleoprotein	Influenza A virus	80	181
16833	FLPSDFPFSV	Capsid protein	Hepatitis B virus	78	236
61151	SSELENFRAYV	Polymerase acidic protein	Influenza A virus	76	246
17516	FQFPQNGQFI	Nucleoprotein	Lymphocytic choriomeningitis virus	75	198
16878	FLRGRAYGL	Epstein-Barr nuclear antigen 3	Human herpesvirus 4 (Epstein Barr virus)	74	166

Figure 2.4 Epitope Detail Search input screen

**Epitope** ?

Epitope ID

Structure Type - Any Epitopes

Organism

Antigen Name

**Epitope Reference Details**

Epitope Structure Defines

Evidence Code for Source...

Epitope Name

Reference Start Position  to

Reference End Position  to

Reference Region

Comments

Data Location in Reference

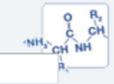
**Epitope Related Object**

Related Object

Type - Any Type

Organism

Antigen Name



**Figure 2.5 Epitope panel on the Specialized Search page expanded to show input fields**

**Epitope** ? 

Epitope ID

Structure Type - Linear Epitopes ▼

Linear Sequence

Match

Modified Residue(s)

Modification(s)

Starting Position  to

Ending Position  to

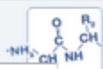
Organism

Antigen Name

Epitope Reference Details

Epitope Related Object

**Figure 2.6 Epitope query fields in the Advanced Search for Linear peptide epitopes**

**Epitope** ? 

Epitope ID

Structure Type - Discontinuous Epitopes ▼

Discontinuous Residues

Modified Residue(s)

Modification(s)

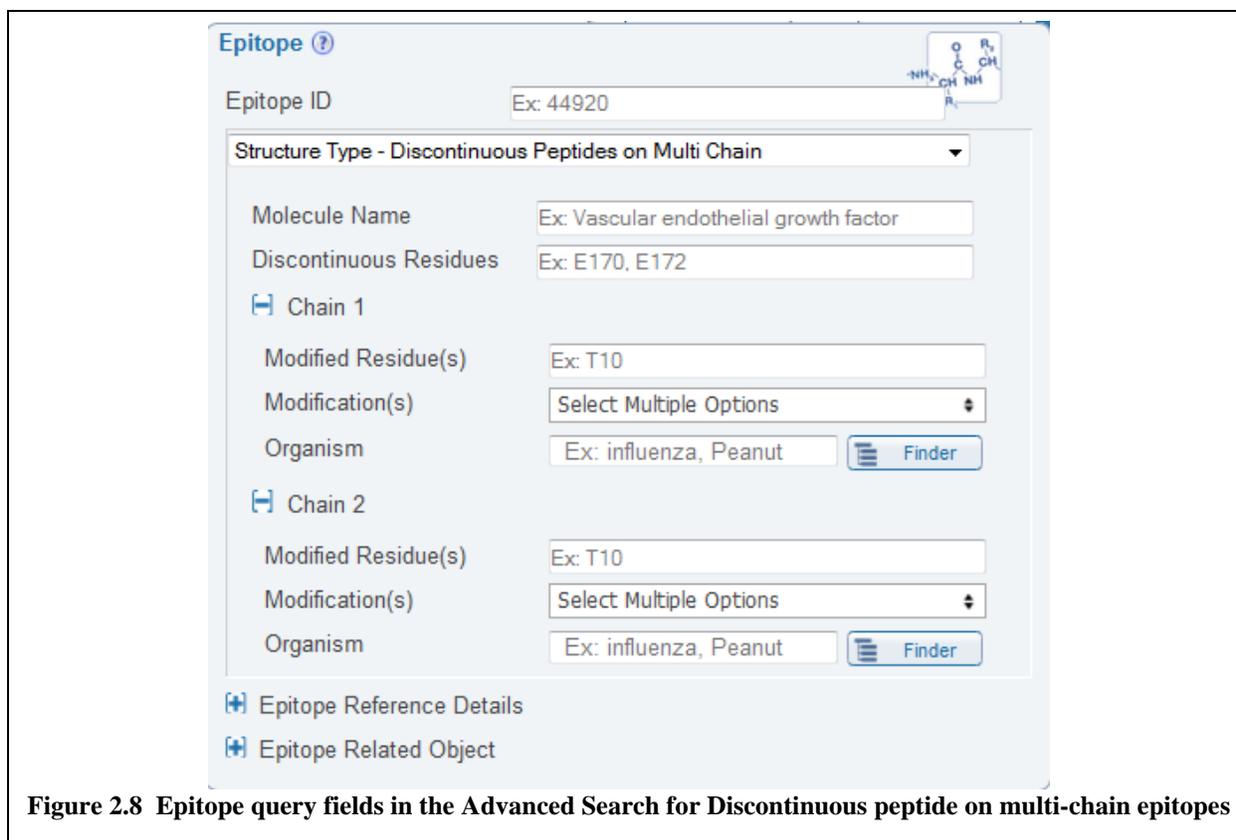
Organism

Antigen Name

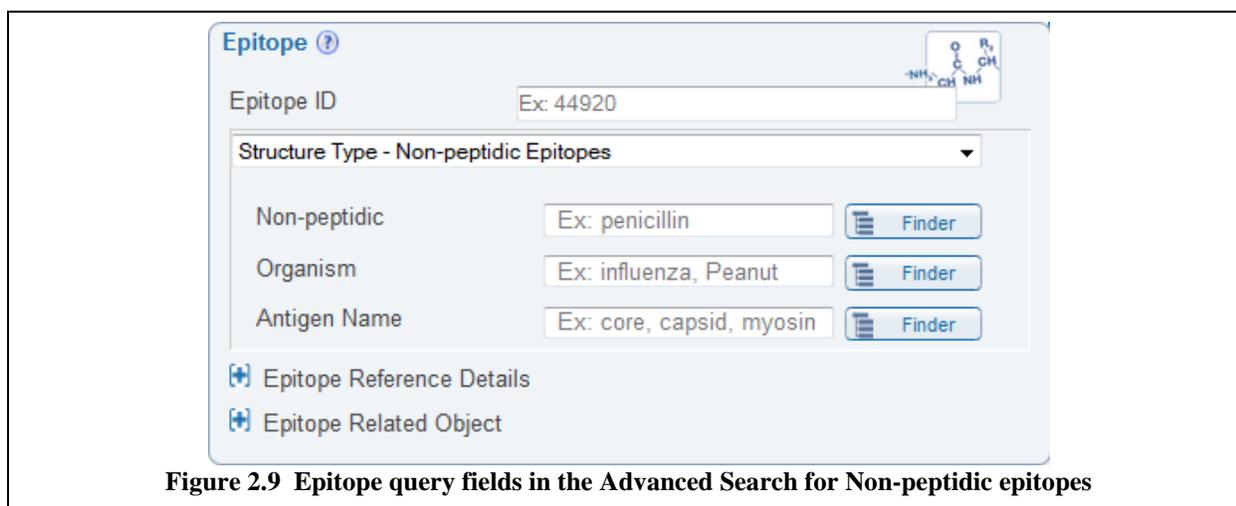
Epitope Reference Details

Epitope Related Object

**Figure 2.7 Epitope query fields in the Advanced Search for Discontinuous peptide epitopes**



**Figure 2.8 Epitope query fields in the Advanced Search for Discontinuous peptide on multi-chain epitopes**



**Figure 2.9 Epitope query fields in the Advanced Search for Non-peptidic epitopes**

There are four other Specialized queries that can be used to select search criteria for references, epitopes, and assays. These are T Cell Assay Details, B Cell Assay Details, MHC Assay Details, and Identifier Search. As an example, the MHC Assay Detailed Search page is shown in Figure 2.10. As is the case for all three assay search pages, the Epitope and Reference criteria fields are listed first and last, respectively, with the other relevant filter criteria specific for each assay type listed between them.

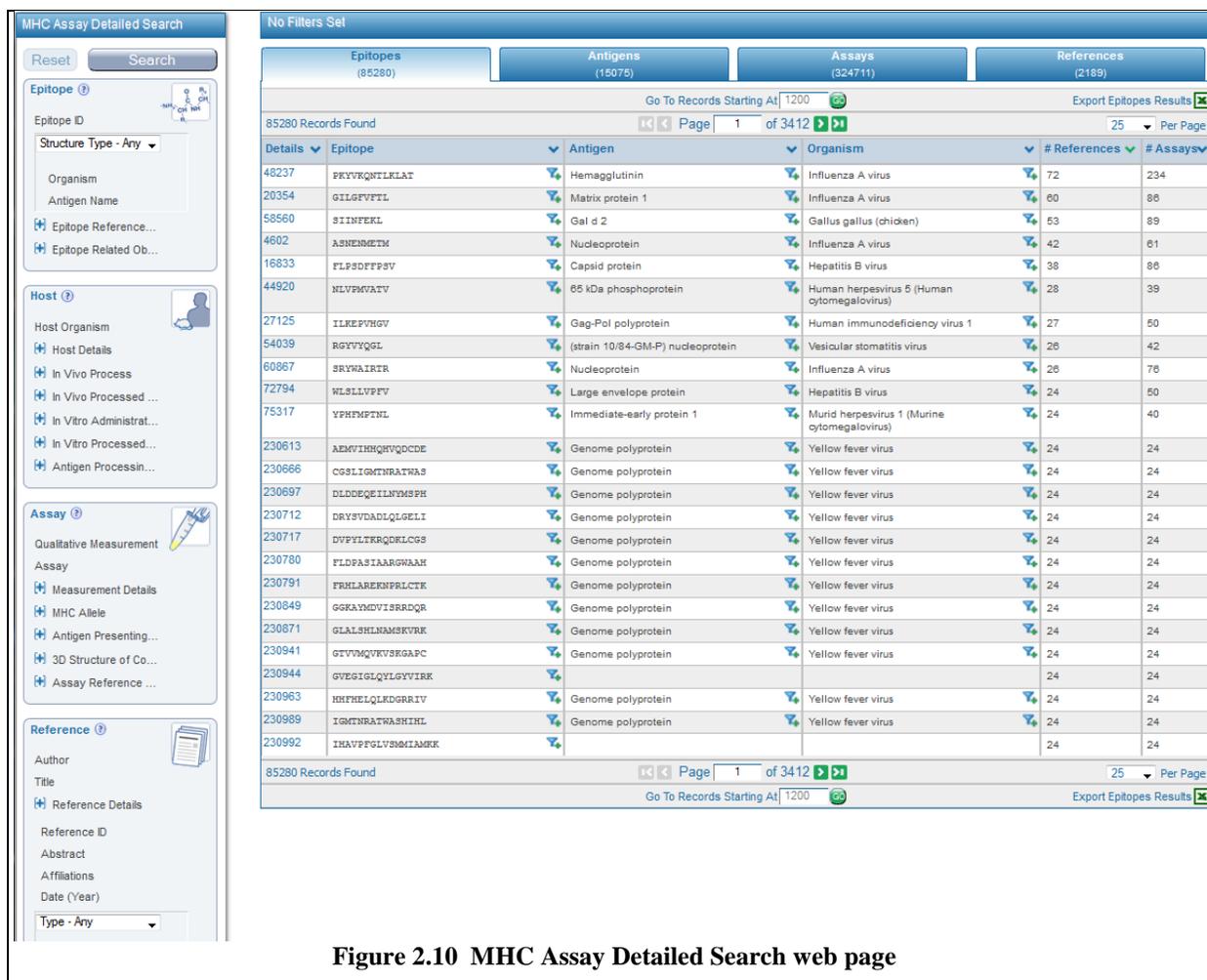


Figure 2.10 MHC Assay Detailed Search web page

The B cell response assay category captures B cell-mediated immunity information and describes antibody responses related to the epitope/antigen. In a B cell response, B lymphocytes (a type of white blood cell) produce proteins called antibodies that bind to antigens. Antibodies are present on the surface of B lymphocytes and are also secreted. Once an antibody binds an antigen, the bound antigen molecule can be engulfed by phagocytes and broken into fragments.

The T Cell Response assay category captures T cell-mediated immunity information where the MHC molecule/antigen complex is recognized by T cells in the context of presentation by an antigen presenting cell.

The MHC Ligand assay category combines assays that were previously in the MHC Binding and MHC Ligand Elution assay categories. As such, the MHC Ligand category captures details relating to the in vitro interaction of the epitope with specific MHC molecules along with available Epitope-MHC complex structure details, that is the epitope's binding capacity to the MHC molecule. It also captures data related to epitopes that are naturally processed and presented on the surface of an antigen presenting cell. The MHC Ligand Elution subcategory differs from the MHC Binding subcategory in that for the former, antigen bound to the MHC molecule on the cell surface has been taken up and processed internally for presentation by the antigen presenting cell, where as for the latter, antigen in solution is bound to MHC molecules on the surface of the antigen presenting cells.

**IDENTIFIER SEARCH**

**IEDB Identifiers**

Epitope ID

Reference ID

Submission ID

Assay ID

**External Identifiers**

PubMed ID

PDB ID

ChEBI ID

**Search**

**Figure 2.11 Identifier Search input screen**

The Identifier Search input screen is shown in Figure 2.11. If users know the epitope, reference, submission, or assay identifier they are seeking, they can enter that information into one of the four IEDB Identifiers fields. Users can also look for information based on the PubMed ID of a reference, the PDB ID of a structure, or the ChEBI ID of a nonpeptidic ligand by using the External Identifiers fields.

### 2.2.3 Search Results Page

The results of a search or query appear on the Results page ([http://www.iedb.org/result\\_v3](http://www.iedb.org/result_v3)). An example is shown in Figure 2.12. The search criteria are listed in the Current Filters section above the results table. There are four tabs for the results. Figure 2.12 displays the Epitopes tab. Results are initially sorted by the number of assays, but the user can click on the column header to change the sort order. A second click will reverse the sort order of the selected column. The table includes the epitope structure description, the corresponding antigen, source organism, the number of references in which the epitope appears, and the number of assays (positive and negative) that reside in the database. Clicking on the Epitope ID number in the far left column will take the user to a Distinct Epitope detail page, as seen in Figure 2.13. The bottom section of the figure contains links to relevant tools in the Analysis Resource. Links can also appear to the Bioinformatics Resource Centers (BRC), such as the Influenza Research Database (IRD). The user can also refine the search results by clicking on the filter icon in the epitope, antigen, and organism fields.

**Pending Filters**

Reset Search

**Epitope** ?

- Any Epitopes
- Linear Epitope
- Discontinuous Epitopes
- Non-peptidic Epitopes

**Antigen** ?

Organism  
Ex: influenza, peanut

Antigen Name  
Ex: core, capsid, myosin

**Assay** ?

- Positive Assays Only
- T Cell Assays
- B Cell Assays
- MHC Ligand Assays

**MHC Restriction** ?

- Any MHC Restriction

**Current Filters:** Positive Assays Only

Epitopes (136124)	Antigens (18043)	Assays (401112)	References (16841)		
136124 Records Found					
Go To Records Starting At 1200					
Page 1 of 5445					
Export Epitopes Results					
Details	Epitope	Antigen	Organism	# References	# Assays
123885	cardiolipin			318	1042
44920	NLVPMVAIV	65 kDa phosphoprotein	Human herpesvirus 5 (Human cytomegalovirus)	223	549
113645	MEVGWYRSPPFSRVVHLYRNGK	Myelin-oligodendrocyte glycoprotein	Mus musculus (mouse)	174	905
20354	GILGFVFTL	Matrix protein 1	Influenza A virus	163	369
58560	SIINFEKL	Gal d 2	Gallus gallus (chicken)	141	346
112741	2,4-dinitrophenyl group			138	474
4602	ASNENMETM	Nucleoprotein	Influenza A virus	129	352
24786	HSLGKWLGHDPKF	Myelin proteolipid protein	Mus musculus (mouse)	111	685
130649	alpha-Gal epitope	Envelope glycoprotein	Murine leukemia virus	107	418
112742	2,4,6-trinitrophenyl group			105	312
130694	1-O-(alpha-D-galactosyl)-N-hexacosanoylphosphatidylcholine			103	424
20788	GLCTLVAML	mRNA export factor ICP27 homolog	Human herpesvirus 4 (Epstein Barr virus)	100	219
6435	CINGVCWIV	Genome polyprotein	Hepatitis C virus	94	237
48237	PKYVQNTLKLAT	Hemagglutinin	Influenza A virus	91	305
61086	SSIEFARL	Envelope glycoprotein B	Human herpesvirus 1 (Herpes simplex virus type 1)	86	285

Figure 2.12 The Epitope tab of the search results page

Distinct Epitope	
Epitope ID	20354
Linear Sequence	GILGFVFTL
Source Organism	Influenza A virus
Source Antigen	Matrix protein 1

Related Information	
References	188 (Click for related results)
Bcell Assays	5 (Click for related results)
Tcell Assays	361 (Click for related results)
MHC Ligand Assays	98 (Click for related results)

Resource	Link
 IEDB-AR: MHC-I Processing <a href="#">↗</a>	Predict MHC class I processing <a href="#">↗</a>
 IEDB-AR: MHC-I <a href="#">↗</a>	Predict MHC class I binding affinity <a href="#">↗</a>
 IEDB-AR: B cell scales <a href="#">↗</a>	Predict B cell epitopes <a href="#">↗</a>
 IRD <a href="#">↗</a>	View sequence feature Influenza A_M1_experimentally-determined-epitope_58(9) details at IRD <a href="#">↗</a>
 IRD <a href="#">↗</a>	View sequence feature Influenza A_M1_monoclonal-antibody-binding-site_58(9) details at IRD <a href="#">↗</a>

**Figure 2.13 An example of the epitope detail page**

The second tab on the results page is the Antigens tab (Figure 2.14). Results are initially sorted by the number of assays, but the user can click on the column header to change the sort order. The table includes the antigen name, the corresponding source organism, the number of associated antigen epitopes in the IEDB, and the number of assays (positive and negative) that reside in the database. The user can explore the data related to a particular antigen by clicking on the funnel icon in the righthand side of the Antigen column. The data in the other three tabs will change accordingly as seen in Figure 2.15.

Epitopes (136124)		Antigens (18043)		Assays (401112)		References (16841)	
Go To Records Starting At 1200 <input type="button" value="GO"/>				Export Antigens Results <input type="button" value="X"/>			
18043 Records Found				Page 1 of 722		25 Per Page	
Antigen	Organism	# Epitopes	# Assays	# References			
Genome polyprotein	Hepatitis C virus	4145	11805	540			
Hemagglutinin	Influenza A virus	1538	5108	531			
Nucleoprotein	Influenza A virus	500	2522	420			
Large envelope protein	Hepatitis B virus	640	2704	307			
Other Homo sapiens (human) protein	Homo sapiens (human)	1478	2304	302			
65 kDa phosphoprotein	Human herpesvirus 5 (Human cytomegalovirus)	313	1732	281			
Matrix protein 1	Influenza A virus	275	1266	276			
Amyloid beta A4 protein	Homo sapiens (human)	195	1583	262			
Gal d 2	Gallus gallus (chicken)	106	718	232			
Circumsporozoite (CS) protein	Plasmodium falciparum (malaria parasite P. falciparum)	357	1563	207			
Myelin basic protein	Homo sapiens (human)	367	1908	206			
Genome polyprotein	Dengue virus	6121	14122	205			
Protein E7	Alphapapillomavirus 9	212	1046	196			
Myelin-oligodendrocyte glycoprotein	Mus musculus (mouse)	69	1107	188			
Pre-glycoprotein polyprotein GP complex	Lymphocytic choriomeningitis mammarenavirus	330	1670	174			
Envelope glycoprotein gp160	Human immunodeficiency virus 1	244	902	166			
Epstein-Barr nuclear antigen 3	Human herpesvirus 4 (Epstein Barr virus)	58	535	161			

Figure 2.14 The Antigens tab of the search results page

Epitopes (500)		Antigens (1)		Assays (2522)		References (420)	
Current Filters: <input checked="" type="checkbox"/> Positive Assays Only <input checked="" type="checkbox"/> Antigen: Nucleoprotein							
Go To Records Starting At 1200 <input type="button" value="GO"/>				Export Antigens Results <input type="button" value="X"/>			
1 Records Found				Page 1 of 1		25 Per Page	
Antigen	Organism	# Epitopes	# Assays	# References			
Nucleoprotein	Influenza A virus	500	2522	420			
1 Records Found				Page 1 of 1		25 Per Page	
Go To Records Starting At 1200 <input type="button" value="GO"/>				Export Antigens Results <input type="button" value="X"/>			

Figure 2.15 The Results filtered by the Nucleoprotein of the Influenza A virus. This was achieved by clicking on the filter icon in the antigen column. The content of the other tabs contain data specific to this antigen.

The third tab on the results page is the Assays tab (Figure 2.16). Results are initially sorted by the Assay Description, but the user can click on the column header to change the sort order. The table contains three additional tabs, one each for the three assay categories – T cell, B cell, and MHC Ligand. The table for the T cell assays includes the assay ID, the reference, epitope, host organism, immunization, assay antigen, antigen epitope relation, MHC restriction, and assay description. The content of the B cell assay tab is the same except it omits the MHC restriction column. The MHC ligand assay tab contains fields

for assay ID, reference, epitope, antigen processing, MHC restriction, assay description, and quantitative measure. Clicking on the Assay ID in the far left lane will take the user to an Assay Detail page. An example is given in Figure 2.17. The actual page contains additional fields below what is shown in the figure, including sections for Epitope Reference Details, Immunization, Host Details, 1<sup>st</sup> In Vivo Process, Administration Details, 1<sup>st</sup> Immunogen, Immunogen Details, Immunization Comments, T Cell Assay, Effector Cells, Antigen Presenting Cells, MHC Allele, Antigen, Antigen Details, and Assay Reference Details. Hyperlinks were added to several fields in 2016: Reference ID, PubMed ID, Epitope ID, Source Accession, Source Organism ID, and MHC Allele Name. The user can also refine the results further by clicking on the filter icon that is available in several of the fields, such as reference, epitope, and host.

Epitopes (136124)		Antigens (18043)		Assays (401112)		References (16841)		
T Cell Assays (109813)		B Cell Assays (104181)		MHC Ligand Assays (187138)				
Go To Records Starting At   A,b  <a href="#">Export T Cell Assays Results</a> 								
109813 Records Found <span style="float: right;">25 Per Page</span>								
Page 1 of 4393								
ID	Reference	Epitope	Host	Immunization	Assay Antigen	Antigen Epitope Relation	MHC Restriction	Assay Description
1244283	H M Vordermeier : Immunology 1993	LFAAFPSFA GLRPTF DTRLM 14 kDa antigen (21-40) Mycobacterium tuberculosis	Mus musculus C57BL/10	Administration in vivo with Mycobacterium tuberculosis H37Ra (Taxonomic Child)	LFAAFPSFAG LRPTF DTRLM 14 kDa antigen (21-40) Mycobacterium tuberculosis	Epitope	H-2-b class II	3H-thymidine proliferation Positive-High
1244288	H M Vordermeier : Immunology 1993	RDGQLTIKA ERTEQK DFDGRS 14 kDa antigen (71-91) Mycobacterium tuberculosis	Mus musculus C57BL/10	Administration in vivo with Mycobacterium tuberculosis H37Ra (Taxonomic Child)	RDGQLTIKA ERTEQK DFDGRS 14 kDa antigen (71-91) Mycobacterium tuberculosis	Epitope	H-2-b class II	3H-thymidine proliferation Positive-High
1481149	T Collen; J Immunol 1991	ETQIQRRQ HTDVSI MDRFVVPN LRGDLQV LAQKVART LP	Bos taurus Friesian	Administration in vivo with Foot-and-mouth disease virus (strain O1) (O1BFS) (Structurally Related) followed by restimulation in vitro	ETQIQRRQH TDVSI MDRFVVPNL RGDLQV LAQKVARTL P	Epitope	class II	3H-thymidine proliferation Positive-High

Figure 2.16 The Assays tab of the search results page. The page contains three additional tabs for T Cell, B Cell, and MHC Ligand Assays

Reference	
Article Authors	H M Vordermeier; D P Harris; R Lathigra; E Roman; C Moreno; J Ivanyi
Article Title	Recognition of peptide epitopes of the 16,000 MW antigen of Mycobacterium tuberculosis by murine T cells.

Reference Detail	
Reference ID	1001056
Abstract	The T-cell repertoire to a prominent immunogen of Mycobacterium tuberculosis has been investigated on the assumption that differences in epitope specificity could influence the protective and pathogenic host reactions. Proliferative responses of lymph node and spleen cells to overlapping peptides, spanning the entire sequence of the 16,000 MW protein antigen were analysed in C57BL/10 and B10.BR mice. Following footpad priming and in vitro challenge with homologous peptide, 12 out of the 14 peptides tested were found to be immunogenic. However, only two peptides of residues 31-40 and 71-91 stimulated strong proliferative responses of T cells from mice which had been presensitized with either killed or live M. tuberculosis organisms; another three peptides were only weakly stimulatory. These epitopes have been immunodominant in both H-2b and H-2k mouse strains, indicating the genetically permissive nature of their recognition. Furthermore, both major immunodominant epitopes were found to be species specific for the M. tuberculosis complex and therefore potentially suitable for the early diagnosis of tuberculosis infection.
Date	1993
Reference Type	Literature
PubMed ID	7503946 <a href="#">🔗</a>
Journal	Immunology
Journal Volume	80
Article Pages	6-12
Journal ISSN	1365-2567
Article Chemical List	Antibodies, Bacterial;Antigens, Bacterial;Epitopes;Peptides
Article MeSH List	Amino Acid Sequence; Animals; Antibodies, Bacterial(biosynthesis); Antigens, Bacterial(immunology); Epitopes(immunology); Lymph Nodes(cytology); Mice; Mice, Inbred C57BL; Molecular Sequence Data; Molecular Weight; Mycobacterium(immunology); Mycobacterium tuberculosis(immunology); Peptides(immunology); Species Specificity; T-Lymphocytes(immunology)

Epitope	
Epitope ID	35787
Chemical Type	Linear peptide
Linear Sequence	LFAAFPSFAGLRPTFDTRLM
Starting Position	21
Ending Position	40
Source Molecule Name	14 kDa antigen
Source Accession	P0A5B7.2 <a href="#">🔗</a>
Source Organism ID	1773
Source Organism	Mycobacterium tuberculosis

**Figure 2.17** The top three sections of a sample T cell assay detail page.

The fourth tab on the results page is the References tab (Figure 2.18). Results are initially sorted by the year of publication or submission, the most recent references first. Again, the user can click on the column header to change the sort order. The table includes the IEDB-assigned reference ID, the PubMed ID, authors, article title, abstract, and the date. For journal articles, the information is downloaded directly from PubMed. Clicking on the Ref ID in the far left lane will take the user to a Reference Detail page. An example is given in Figure 2.19.

Go To Records Starting At 1200 						Export References Results 
16841 Records Found						Page 1 of 674 
Ref ID	PMID	Author	Title	Abstract	Date	
1028722 	25642819 	Richard W Birkinshaw; Daniel G Pellicci; Tan-Yun Cheng; Andrew N Keller; Maria Sandoval-Romero; Stephanie Gras; Annemieke de Jong; Adam P Uldrich; D Branch Moody; Dale I Godfrey; Jamie Rossjohn	$\alpha\beta$ T cell antigen receptor recognition of CD1a presenting self lipid ligands.	A central paradigm in T cell-mediated immunity is the simultaneous co-recognition of antigens and antigen-presenting molecules by the T cell antigen receptor (TCR). CD1a presents a broad repertoire ... <a href="#">more...</a>	2015	
1028867 	25185583 	D Goedhals; J T Paweska; F J Burt	Identification of human linear B-cell epitope sites on the envelope glycoproteins of Crimean-Congo haemorrhagic fever virus.	A peptide library was used to screen for regions containing potential linear B-cell epitope sites in the glycoproteins and nucleoprotein of Crimean-Congo haemorrhagic fever virus (CCHFV) in an enzyme- ... <a href="#">more...</a>	2015	
1028697 	25668439 	Hongquan Wan; Hua Yang; David A Shore; Rebecca J Garten; Laura Couzens; Jin Gao; Lianlian Jiang; Paul J Carney; Julie Villanueva; James Stevens; Maryna C Eichelberger	Structural characterization of a protective epitope spanning A(H1N1)pdm09 influenza virus neuraminidase monomers.	A(H1N1)pdm09 influenza A viruses predominated in the 2013-2014 USA influenza season, and although most of these viruses remain sensitive to Food and Drug Administration-approved neuraminidase (NA) inh ... <a href="#">more...</a>	2015	
1028889 	25855295 	Hiroaki Tanabe; Yoshifumi Fujii; Miki Okada-Iwabu; Masato Iwabu; Yoshihiro Nakamura; Toshiaki Hosaka; Kanna Motoyama; Mariko Ikeda; Motoaki Wakiyama; Takaho Terada; Noboru Ohsawa; Masakatsu Hato; Satoshi Ogasawara; Tomoya Hino; Takeshi Murata; So Iwata; Kunio Hirata; Yoshiaki Kawano; Masaki Yamamoto; Tomomi Kimura-Someya; Mikako Shirouzu; Toshimasa Yamauchi; Takashi Kadowaki; Shige-yuki Yokoyama	Crystal structures of the human adiponectin receptors.	Adiponectin stimulation of its receptors, AdipoR1 and AdipoR2, increases the activities of 5' AMP-activated protein kinase (AMPK) and peroxisome proliferator-activated receptor (PPAR), respectively, t ... <a href="#">more...</a>	2015	

Figure 2.18 The References tab of the search results page

Reference	
Article Authors	D Goedhals; J T Paweska; F J Burt
Article Title	Identification of human linear B-cell epitope sites on the envelope glycoproteins of Crimean-Congo haemorrhagic fever virus.

Reference Detail	
Reference ID	1028867
Abstract	A peptide library was used to screen for regions containing potential linear B-cell epitope sites in the glycoproteins and nucleoprotein of Crimean-Congo haemorrhagic fever virus (CCHFV) in an enzyme-linked immunosorbent assay (ELISA). The library consisted of 156 peptides, spanning the nucleoprotein and mature GN and GC proteins in a 19-mer with 9-mer overlap format. Using pooled serum samples from convalescent patients to screen the library, six peptides were identified as potential epitope sites. Further testing of these six peptides with individual patient sera identified two of these peptides as probable epitope sites, with peptide G1451-1469 reacting to 13/15 and peptide G1613-1631 to 14/15 human sera. These peptides are situated on the GC protein at amino acid positions 1451-1469 (relative to CCHFV isolate SPU103/97) (CTGTCYACSSGISCKVRIH) and 1613-1631 (FMFGWRILFCFKCRRTRG). Identified peptides may have application in ELISA for diagnostic or serosurveillance purposes.
Affiliations	Department of Medical Microbiology and Virology, National Health Laboratory Service/University of the Free State, Bloemfontein, South Africa; Center for Emerging and Zoonotic Pathogens, National Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg, South Africa.
Date	2015
Reference Type	Literature
PubMed ID	25185583
Journal	Epidemiol Infect
Journal Volume	143
Article Pages	1451-6
Journal ISSN	1469-4409
Curation Last Updated	2015-04-23 20:01:17

Related Information	
Epitopes	54 ( <a href="#">Click for related results</a> )
Bcell Assays	54 ( <a href="#">Click for related results</a> )
Tcell Assays	0
MHC Ligand Assays	0

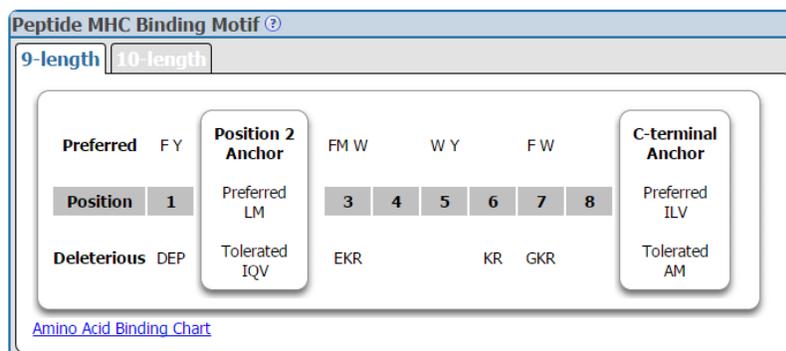
Figure 2.19 An example of the Reference detail page

### 2.2.3.1 Peptide MHC Binding Motif Displays

A display of peptide MHC binding motif diagrams and amino acid binding charts for a subset of MHC allele records appears on the MHC Allele Details page. Clicking on the allele name in the MHC Restriction column on the Assays tab for T Cells and MHC Ligand will bring the user to an MHC Allele Details page (Figure 2.20). As can be seen by the tabs in the binding motif display, the Peptide MHC Binding Motif diagrams on MHC Allele Information pages support diagrams for multiple lengths.

MHC Allele	
Allele Name	HLA-A*02:01
Synonyms	HLA-A2.1,HLA-A*0201
Includes	HLA-A*020101
Restriction Level	complete molecule
Organism	human (Homo sapiens)
Class	I
Locus	A
Serotype	A2
Molecule	HLA-A*02:01
Chain 1 Name	HLA-A*02:01
Chain 1 Accession	HLA00005 <a href="#">↗</a>
Chain 2 Name	Beta-2-microglobulin
Chain 2 Accession	48428791 <a href="#">↗</a>

Related Information	
Epitopes	24921 (Click for related results)
Bcell Assays	0
Tcell Assays	12130 (Click for related results)
MHC Ligand Assays	39116 (Click for related results)



**Figure 2.20** Sample MHC Allele Details page with Peptide MHC Binding Motif diagram for HLA-A\*02:01. Note the link to the corresponding Amino Acid Binding Chart on the bottom left of the diagram

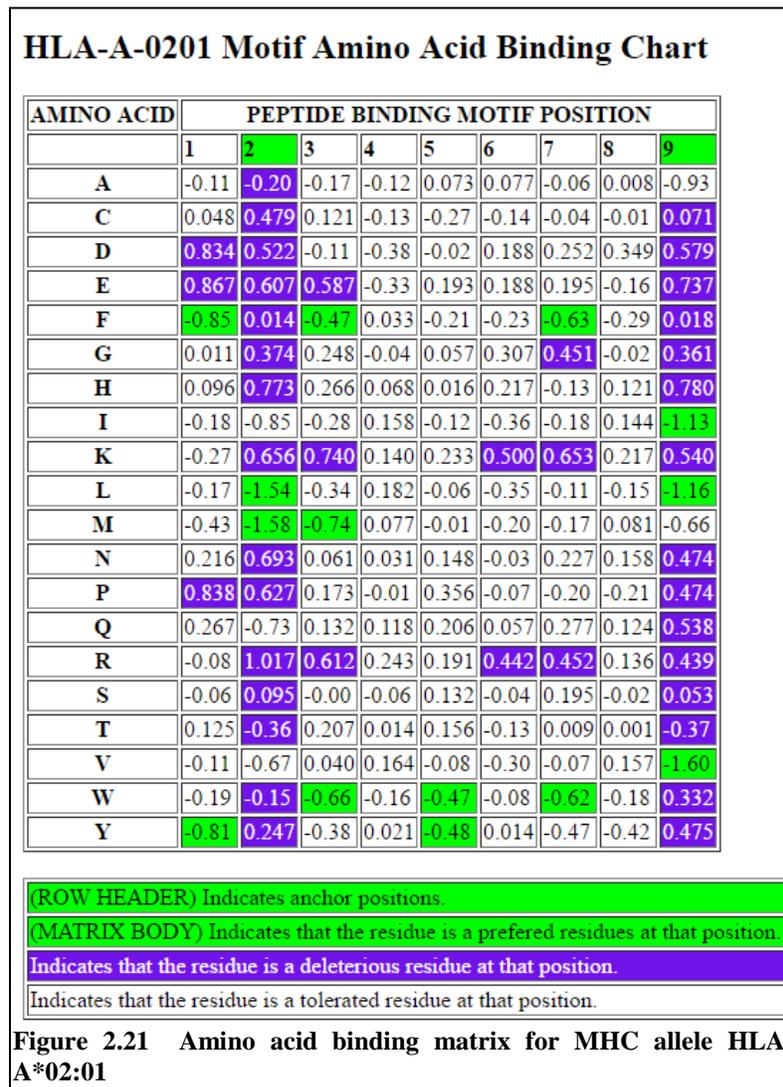
The Peptide MHC Binding Motif diagram summarizes the results of the amino acid binding for each position, which is shown in Figure 2.21. The amino acids are color-coded for each position, where green indicates the residue is preferred for binding, purple indicates it is deleterious for binding, and white indicates the residue is tolerated at that position. The motif diagram shows which residues are preferred and tolerated at the anchor positions, and which residues are preferred and deleterious at the non-anchor positions.

The residue preference is calculated from the SMM prediction matrix, which is described in Section 2.3.1.1. The following steps are taken in determining the residue preference at each position:

For anchor position:

1. determine the best value of this position from corresponding column of the SMM matrix
2. if a residue's value in this column of the SMM matrix is within 3 fold of the best value, designate this residue as preferred
3. if a residue's value in this column of the SMM matrix is within 10 fold of the best value, designate this residue as tolerated

- The remaining residues are designated as deleterious



For non-anchor position:

- determine the median value of this position from corresponding column of the SMM matrix
- if a residue's value in this column of the SMM matrix is within 3 fold of the median value, designate this residue as tolerated
- if a residue's value in this column of the SMM matrix is above 3 fold of the median value, designate this residue as preferred
- if a residue's value in this column of the SMM matrix is below 3 fold of the median value, designate this residue as deleterious

### 2.2.3.2 Immunome Browser

The Immunome Browser is a feature integrated into the IEDB to more efficiently display large sets of immune epitope data. The purpose of the tool is to allow users to explore how often each protein region has been studied in immune assays and in how many assays the immune response was positive or negative. The Immunome Browser “maps” query results from the IEDB onto a reference proteome. A reference proteome is used because (1) epitopes reported in IEDB were identified for different strains and

protein isoforms – mapping to the reference protein allows to visualize and study such epitopes as they would have the same antigen; (2) different mutant variants of the same epitope were tested and reported; and (3) immune response varies among studies and assays due to heterogeneity of samples and complexity of immune response. The Immunome Browser presents the mapped data along with calculated ‘epitope prominence scores’ that rank known epitopes based on how frequently subjects responded to each epitope, allowing identification of immunodominance. Both linear and discontinuous linear peptide epitopes with T cell or B cell responses are supported for viewing using the Immunome Browser. It is accessible via the Immunome Browser icon on the Antigen tab on the results page, as seen in Figure 2.22.



**Figure 2.22** The Immunome Browser is accessed by clicking on the icon shown on the Antigen tab on the results page of a query

In order to make the Immunome Browser capability more prominent, a pop-up window was added that appears the first time a user goes to the Antigen tab after making a search, as seen in Figure 2.23. It can be easily dismissed by clicking on the large X in the upper right-hand corner or clicking anywhere else on the window.



**Figure 2.23** A pop-up window appears each time the user clicks on the Antigen tab on the Results page. It highlights the Immunome Browser icon and briefly explains the Immunome Browser’s functionality.

In the Immunome Browser, each selected antigen is considered to be a target protein, and the tool by default maps and visualizes all epitopes in the query result that are from the source antigen, which is the target protein or any of its taxonomic child. The mapping is done using the alignment between the source antigen and the target protein, and matching the epitope positions in the source antigen with the residue positions in the target protein. By default, epitopes that were tested as negative for the correspondent query will be mapped as well, even if the user queried only for positive assays. Also, the epitopes that failed to satisfy the default mapping criteria won’t be mapped and visualized. The Immunome Browser output for query criteria of T cell assays, linear epitopes, Hepatitis C virus, and human host is shown in Figure 2.24.



**Figure 2.24 Immunome Browser graphical output displaying assay results mapped onto a reference proteome by residue position**

Two graphs are generated in the Immunome Browser. The upper plot titled “Response Frequency” renders the lower (dark color) and upper (light color) bounds of the 95% confidence interval (CI) of the response frequency (RF) for each target protein position, averaged over all epitopes mapped to that position. RF is calculated as the number of positively responded subjects (or individuals in this case) relative to the total number of those tested summed up by mapped epitopes. The graph, however, displays only the upper and lower bounds of the 95% confidence interval (CI) for the RF score and not the RF score itself. The lower plot, titled “Epitope Assay Counts,” indicates the number of positive and negative assays for each residue position.

Results Returned: 3464 | Displaying: 3464  Display Graphed Residue Positions [Export Results](#)

Epitope ID	Epitope Sequence	Mapped Position	Identity	Subjects Tested	Subjects Responded	Assays Positive	Assays Negative	Response Freq.(95% CI)
42691	MSTNFKPQR	1-9	100%	3	0	0	1	0.00 (0.00:0.61)
42698	MSTNFKPQRKTKRNTNRR	1-18	100%	7	1	1	0	0.14 (0.01:0.50)
42673	MSTIIPKQQRKTRN	1-14	92%	60	3	1	0	0.05 (0.02:0.14)
42683	MSTNFKPQKGNKRNT	1-15	86%	2	0	0	2	0.00 (0.00:0.76)
42682	MSTNFKPQK	1-9	88%	4	0	0	2	0.00 (0.00:0.51)
42685	MSTNFKPQKGNKRNTNRRPQ	1-20	90%	23	1	1	1	0.04 (0.00:0.18)
42679	MSTNFKKFRKTRN	1-15	85%	60	5	1	0	0.08 (0.04:0.18)
42696	MSTNFKPQRKTRNT	1-15	100%	1	0	0	1	0.00 (0.00:0.94)
42699	MSTNFKPQRKTKRNTNRRPQ	1-20	100%	6	1	1	2	0.17 (0.01:0.56)
42681	MSTNFKPQIKTKRNTNRR	1-19	94%	1	0	0	1	0.00 (0.00:0.94)

**Figure 2.25 Immunome Browser tabular output.**

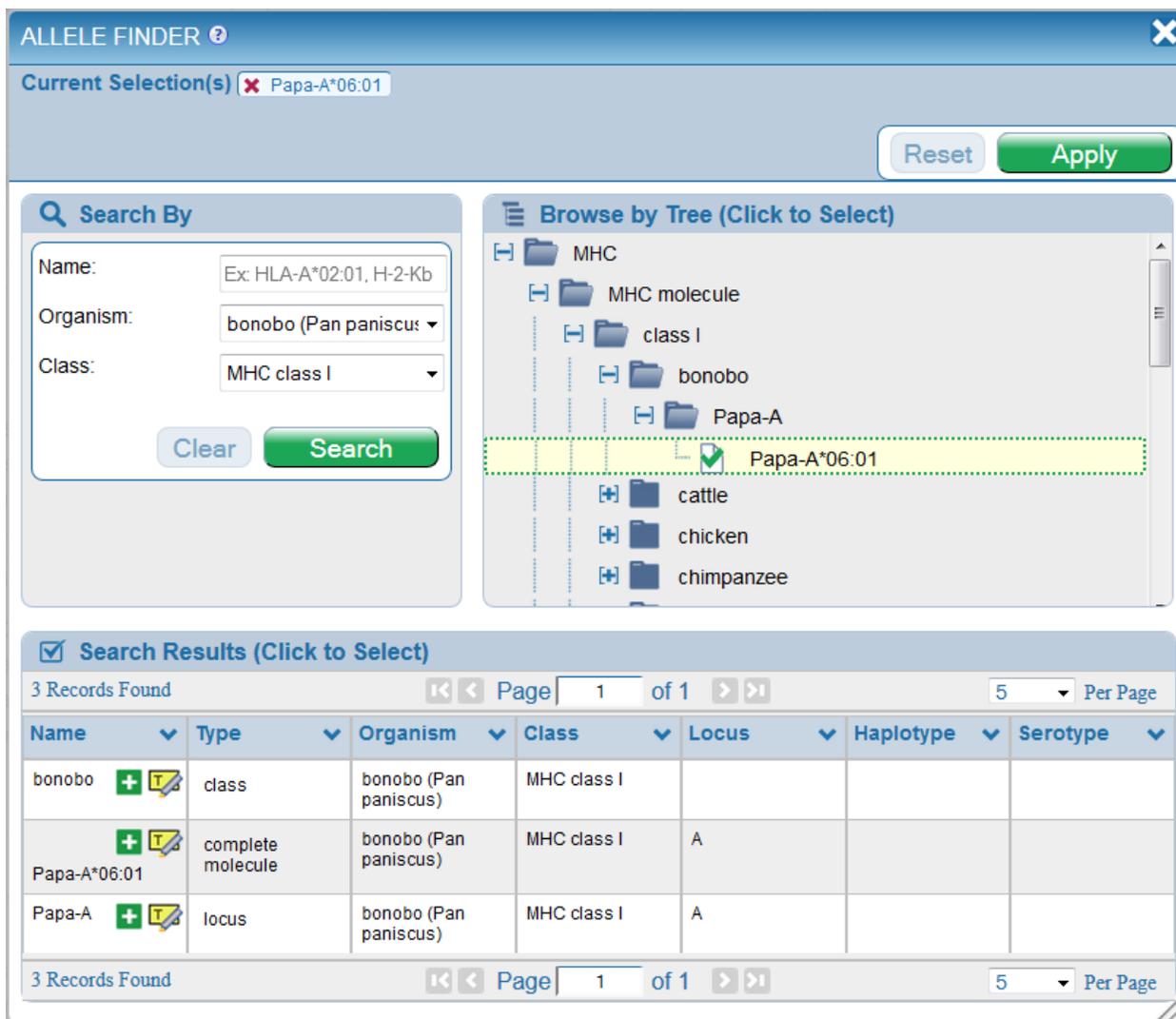
In addition to the plots, a table is generated that displays the epitope ID, sequence, position, and identity, as well as the number of subjects tested, the number who responded, and number of positive and negative assays, and the response frequency (Figure 2.25). The user can interactively zoom in and out on the plots to a specific protein region and, if the check box for the table to correspond the plots is selected, the table is automatically updated to show only epitope from the selected region. Alternatively, the user can specify the residue start and end positions at the bottom of plots and click Update Display.

## **2.2.4 Finders Overview**

Several finders (Allele, Assay, Disease, Molecule, Geolocation, and Organism) are available to help facilitate selections and control vocabulary usage, thus improving query results. At times the potential list of selections can be quite extensive, and the finders help users make selections from large lists. The finders can be utilized when performing all queries. Multiple selections can be made when utilizing finders during a query.

### **2.2.4.1 Allele Finder**

The MHC Allele Finder facilitates the selection of one or more MHC alleles. The finder consists of four sections as shown in Figure 2.26. The top section lists the alleles that the user selects. The selections can be removed by clicking on the red X of an allele name or by clicking on the Reset button to delete all selected alleles. When the user has completed the selection process, the user needs to click the green Apply button. The middle left section of the finder contains the “Search By” panel. Users can search by allele name, source organism, and/or class (I, II, or non-classical). After the user supplies their search criteria and clicks the Search button, the system will filter the list of MHC alleles using the organism, class, and allele provided. The allele finder uses wild card characters by default on both ends of criteria entered into the allele field. The system then returns in the bottom “Search Results” panel a list of any alleles that contain the value in the name field and match the class selected. Users can click on the green “+” icon to select an allele. Users can also click on the yellow highlighter icon to highlight the allele in the allele tree in the middle right section of the finder. Alternatively, a user can browse the hierarchical allele tree. Clicking on a node in this tree will add the allele to the Current Selections section of the finder.



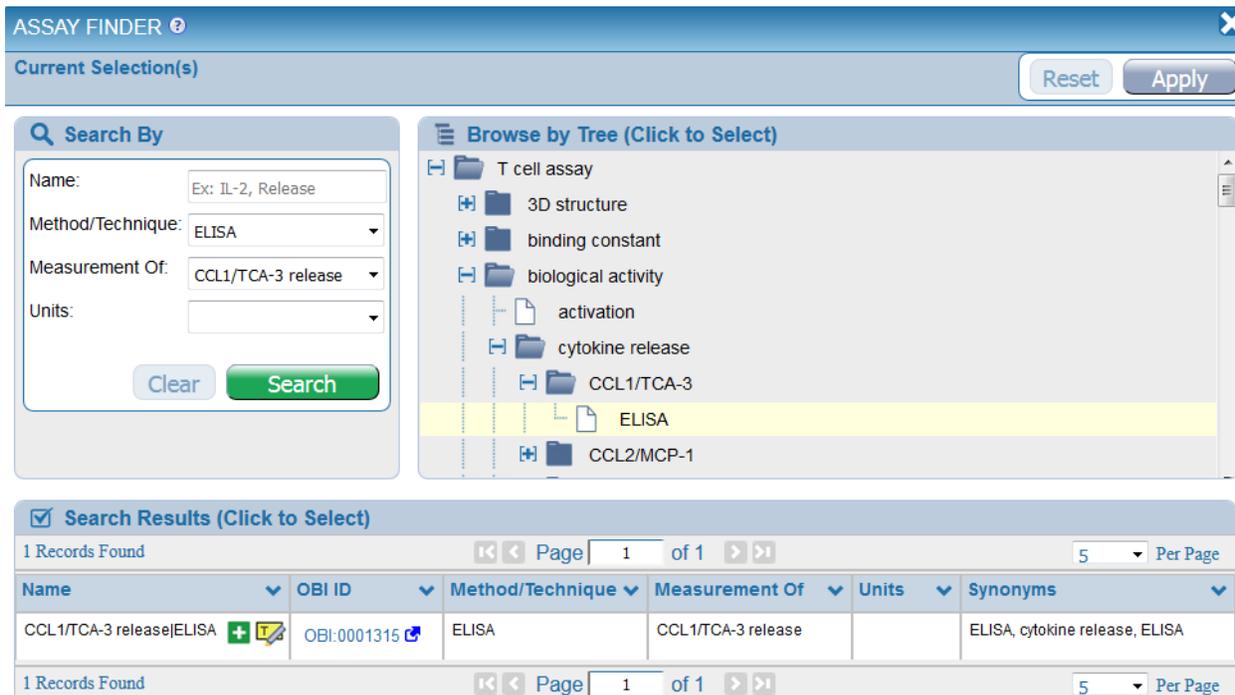
**Figure 2.26 Allele Finder.** Alleles are displayed in a tree structure and users can search by allele name, source organism, and MHC class (I, II, or non-classical)

### 2.2.4.2 Assay Finder

The IEDB contains experimental data characterizing antibody and T cell epitopes studied in humans, non-human primates, and other animal species. Search results may easily be narrowed to include only the specific assay categories of T Cell Assays, B Cell Assays, and MHC Ligand Assays. Additionally, one may further narrow search results to very specific experimental methods using the Assay Finder.

The Assay Finder is used to facilitate the selection of one or more assay types and lists all assay types in the selected assay category. The Assay Finder for T Cell, B cell, and MHC Ligand assays can be found in the Assay section on the the home page search interface, the left-hand side of the Results page, and on their respective Specialized Search pages. These versions of the Assay Finder organize and display assays in a tree structure, as seen in Figure 2.27 for the T Cell assays. Users can also search for assays by their method/technique and by the characters within the assay's name. The figure shows the results for ELISA type assays for "CCL1/TCA-3 release". Users can click on the green "+" icon to select an assay. Users can also click on the yellow highlighter icon to highlight the assay in the assay tree in the middle right section of the finder, as is shown in the figure. Alternatively, a user can browse the hierarchical

assay tree. Clicking on a node in this tree will add the assay to the Current Selections section of the finder.

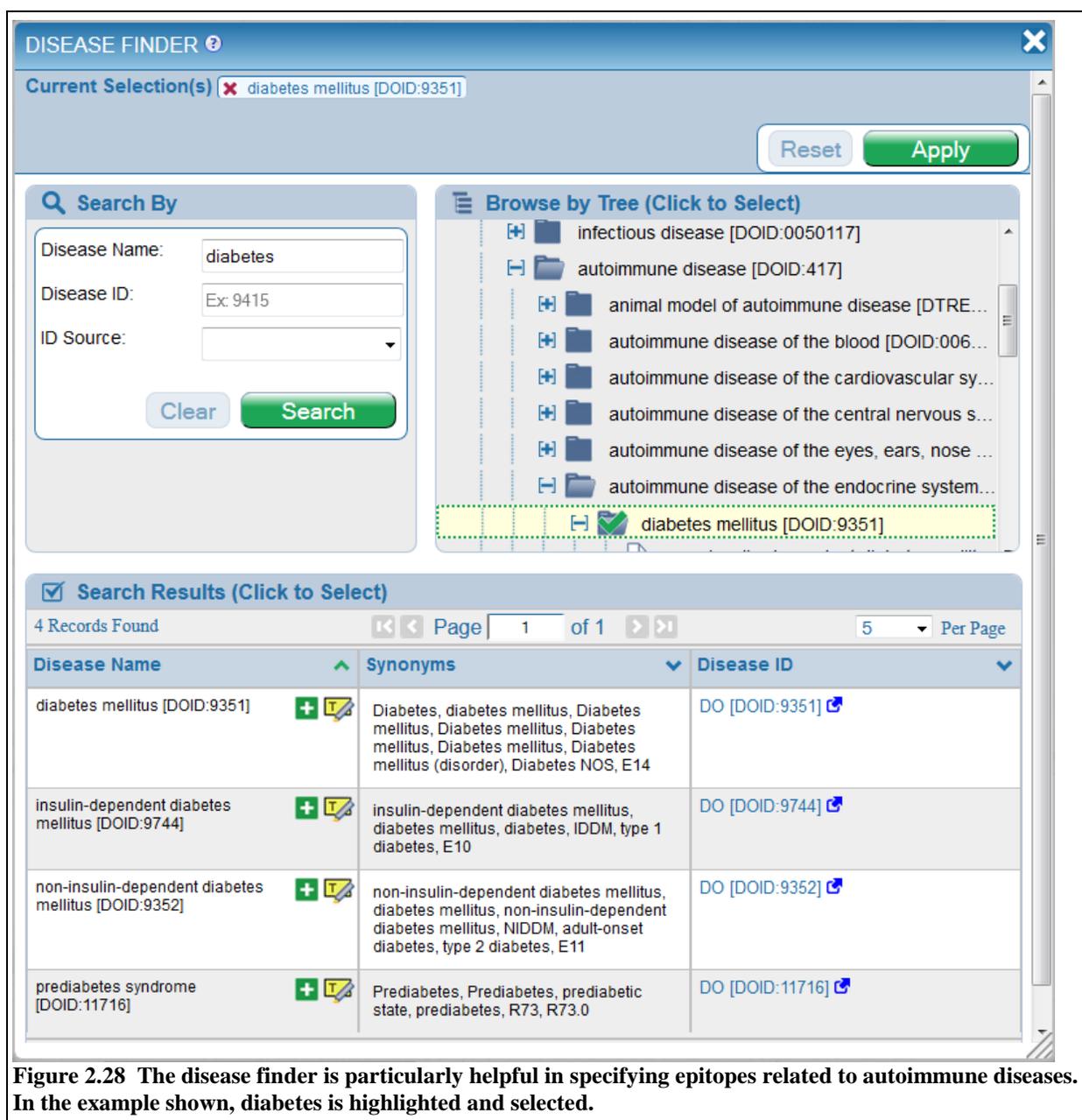


**Figure 2.27 Assay Finder on the T Cell Advanced Search page. Assays are displayed in a tree structure and users can find displays by their method/technique and by the characters with which the assay name starts**

### 2.2.4.3 Disease Finder

This query feature allows users to select a disease association based on the host disease state or status as captured in curation. Diseases are organized into a browsable tree structure. Upon the selection of host disease state/status, the system provides the user with a hierarchical *Disease Tree* that organizes the data at the highest level according to the five disease categories (Allergy, Autoimmunity, Infectious Disease, Additional Diseases by category, and Transplant-related Disease and Allo-reactivity) and then by the anatomical location of the disease (e.g. respiratory tract, gastrointestinal tract). In this way the user can either type in the disease name or synonym of interest, or navigate the Disease Tree by disease manifestation and physical location to see what is available. A node for healthy subjects is also included. The Disease Finder is fully functional for only Autoimmunity. The finder will be revised with an updated disease ontology and hierarchical tree during 2017 based on the Disease Ontology that will make it functional for the other disease categories. Infectious disease searches are best accommodated using the Organism Finder to specify the antigen.

Figure 2.28 shows the Disease Tree with diabetes highlighted. DOID refers to the Disease Ontology identifier. The DTREE designates an IEDB-specific identifier that will be updated later with a DOID number when the Disease Ontology incorporates the terms.



#### 2.2.4.4 Molecule Finder

The Molecule Finder is used to facilitate the selection of source antigens, immunogens, and epitopes. There are two versions of the Molecule Finder – one specifically for non-peptidic molecules found in the Epitope box in the results page, and another that includes peptidic and non-peptidic branches, found in the Antigen box in the results page. The former contains the structures curated by the Chemical Entities of Biological Interest (ChEBI) database. An example of the non-peptidic finder is shown in Figure 2.29.

NON-PEPTIDIC MOLECULE FINDER

Current Selection(s) x penicillin Reset Apply

**Search By**

Name:

Molecule ID:

Clear Search

**Browse by Tree (Click to Select)**

- monobactam
- oxacephem
- penams
  - penamcarboxylate
  - penicillanic acid ester
  - penicillanic acids
    - 6-aminopenicillanic acid
    - 6-formamidopenicillanic acid
    - penicillin

**Search Results (Click to Select)**

19 Records Found Page 1 of 4 5 Per Page

Molecule Name	Synonyms	Database ID	Organism Name
penicillin	penicillins, Penicillin, penicillins, [H][C@]12SC(C)(C)[C@@H](N1C(=O)[C@H]2NC(=O)C(O)=O, C9H11N2O4SR	ChEBI:17334	
penicillinate anion	C9H10N2O4SR, C9H10N2O4S, penicillin anion, penicillin, penicillin anions, [H][C@]12SC(C)(C)[C@@H](N1C(=O)[C@H]2NC(=O)C(O)=O	ChEBI:51356	
benzylpenicillin	Benzylpenicillin, 2,2-dimethyl-6beta-(phenylacetamido)penam-3alpha-carboxylic acid, benzylpenicilline, 6-(2-phenylacetamido)penicillanic acid, bencilpenicilina, InChI=1S/C16H18N2O4S/c1-16(2)12(15(21)2 ...more...	ChEBI:18208	
penicillin O	6beta-[(prop-2-en-1-ylsulfanyl)acetamido]-2,2-dimethylpenam-3alpha-carboxylic acid, allithiomethylpenicillin, (2S,6R,6R,6L)-	ChEBI:51207	

**Figure 2.29** The non-peptidic version of the Molecule Finder. This finder is accessed in the Epitope filter on the search results page.

With regard to the protein or peptidic branch of the Molecule Finder, individual GenPept proteins utilized by IEDB data are assigned to parent proteins from reference proteomes by sequence homology. These reference proteomes are graded by a star system described below that reflects the quality and completeness of each. An example of the star system and the protein tree can be seen in Figure 2.30.

The screenshot shows the MOLECULE FINDER interface. At the top, the current selection is 'Voltage-gated potassium channel' with two stars. Below this, there are two main sections: 'Search By' and 'Browse by Tree (Click to Select)'. The 'Search By' section has fields for Name (potassium channel), Molecule ID (Ex: P69710), and Source Organism (Ex: influenza, peanut). The 'Browse by Tree' section shows a hierarchical view of categories like 'material entity', 'non-peptidic material', and 'protein', with 'Voltage-gated potassium channel' selected. Below these is the 'Search Results (Click to Select)' section, which displays a table of 87 records. The table has columns for Molecule Name, Synonyms, Database ID, and Organism Name. Three rows are visible, each representing a different potassium channel protein from Homo sapiens (human).

Molecule Name	Synonyms	Database ID	Organism Name
ATP-sensitive inward rectifier potassium channel 1 ★★	Potassium channel, KCNJ1_HUMAN, Potassium channel, inwardly rectifying subfamily J member 1, ATP-regulated potassium channel ROM-K, inwardly rectifying subfamily J member 1, ATP-sensitive inward recti ...more...	UniProt [P48048]	Homo sapiens (human)
ATP-sensitive inward rectifier potassium channel 10 ★★	ATP-dependent inwardly rectifying potassium channel Kir4.1, ATP-sensitive inward rectifier potassium channel 10, inward rectifier K+ channel KIR1.2, Potassium channel, inwardly rectifying subfamily J ...more...	UniProt [P78508]	Homo sapiens (human)
ATP-sensitive inward rectifier potassium channel 11 ★★	Potassium channel, inwardly rectifying subfamily J member 11, Potassium channel, inwardly rectifying subfamily J member 11, inward rectifier K(+) channel Kir6.2, KCJ11_HUMAN, IKATP, ATP-sensitive inwa ...more...	UniProt [Q14654]	Homo sapiens (human)

Figure 2.30 An example of the protein branch of the Molecule Finder

## Proteomes

★★★ For some well-studied species UniProt provides reference proteomes that contain a full set of all proteins expressed by the species. For some bacterial species having inconsistent protein expression, additional proteins have been added to the reference proteome to create metaproteomes. These reference proteomes or metaproteomes are designated by **three stars**.

★★ For other species that have been completely sequenced, UniProt provides complete proteomes. In addition, for some species expressing allergens, formal nomenclature designated by the International Union of Immunological Societies (IUIS) exists to describe these allergens. Complete proteomes that are not considered reference proteomes, or ones that contain formal IUIS allergen nomenclature for a subset of proteins, are designated by **two stars**.

★ For some species, a proteome does not currently exist in UniProt, but GenBank provides a set of proteins representative of the species. These GenBank proteomes are designated by a **single star**.

☆ For species that have no proteome in UniProt or GenBank, and no IUIS nomenclature, UniProt may still contain some records that can be used as parents. This case is designated with an **unfilled star**.

**No Star.** Species having no proteome in either UniProt or GenBank are designated by **no stars**.

## Proteins

Within each species' proteome, individual, "parent" proteins serve to group multiple distinct GenPept sequences. These GenPept entries are the "children" for each proteome protein in the Molecule Tree. This allows users to search IEDB data by selecting the parent protein from the reference proteome, rather than having to select each individual GenPept entry. The "parent" proteins within each proteome also use stars to denote the quality of information provided by each.

★★ UniProt reviewed proteins or proteins having official IUIS allergen nomenclature have **two stars**.

★ UniProt unreviewed proteins or proteins from GenBank have a **single star**.

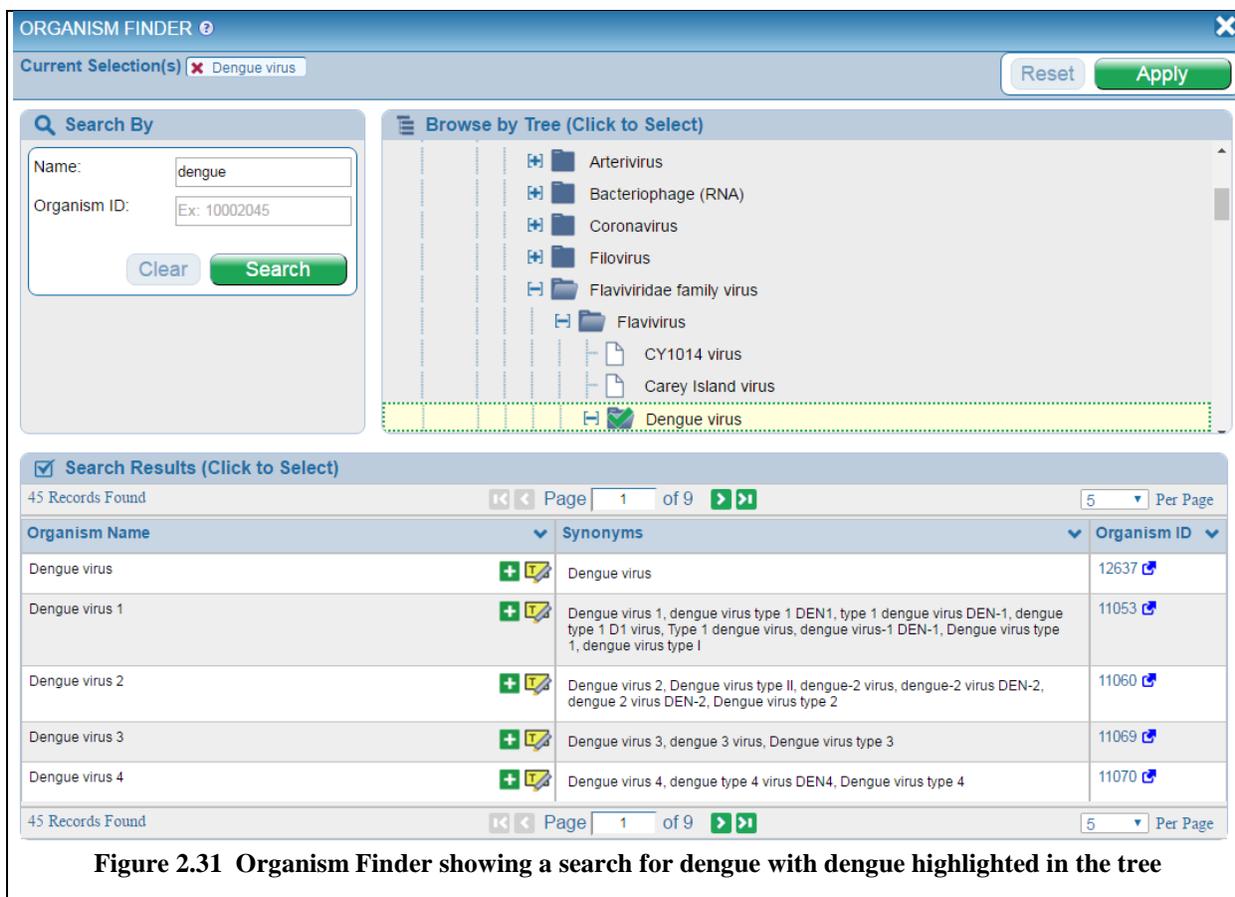
☆ Nodes of the protein branch of the molecule tree containing GenPept and IEDB internal protein accessions having no homology to any protein within a reference proteome are designated with an **unfilled star**.

Organizational nodes, utilized by the Molecule tree to clarify the relationship between groups of similar proteins have **no stars**. An example of these nodes is "Immunoglobulin" used to group all immunoglobulin proteins from a single species.

As one can see in Figure 2.29, a user can search the molecule tree by entering text, including synonyms, in the Name field in the upper lefthand corner of the finder. The user can also specify the source organism of the molecule of interest using the autocomplete field or the organism finder. For example, this can facilitate the specification of a hemagglutinin is a particular strain of Influenza.

#### 2.2.4.5 Organism Finder

The organism finder is used to facilitate the selection of a species or strain from the NCBI Taxonomy Database. The Organism finder allows the user to find species using their name, synonyms, or taxonomy identifier (assigned by NCBI). When the user performs a search, the system will display a tree structure. The user can search for a name or Organism ID in the "Search By" panel in the upper lefthand portion of the Finder. In the example shown in Figure 2.31, a search for "dengue" is performed, which results in 45 items being found. The user can select one or more of the items by clicking on the green "+" icon or clicking a branch in the tree. The user can also choose to highlight the item in the tree, as shown in the figure, by clicking the corresponding yellow highlighter icon. Synonyms for a selection are listed next to the organism name column in the search results table. When the Organism Finder is adjacent to a Host Organism field, only the source organisms for which data exist in the IEDB are displayed in the tree structure.



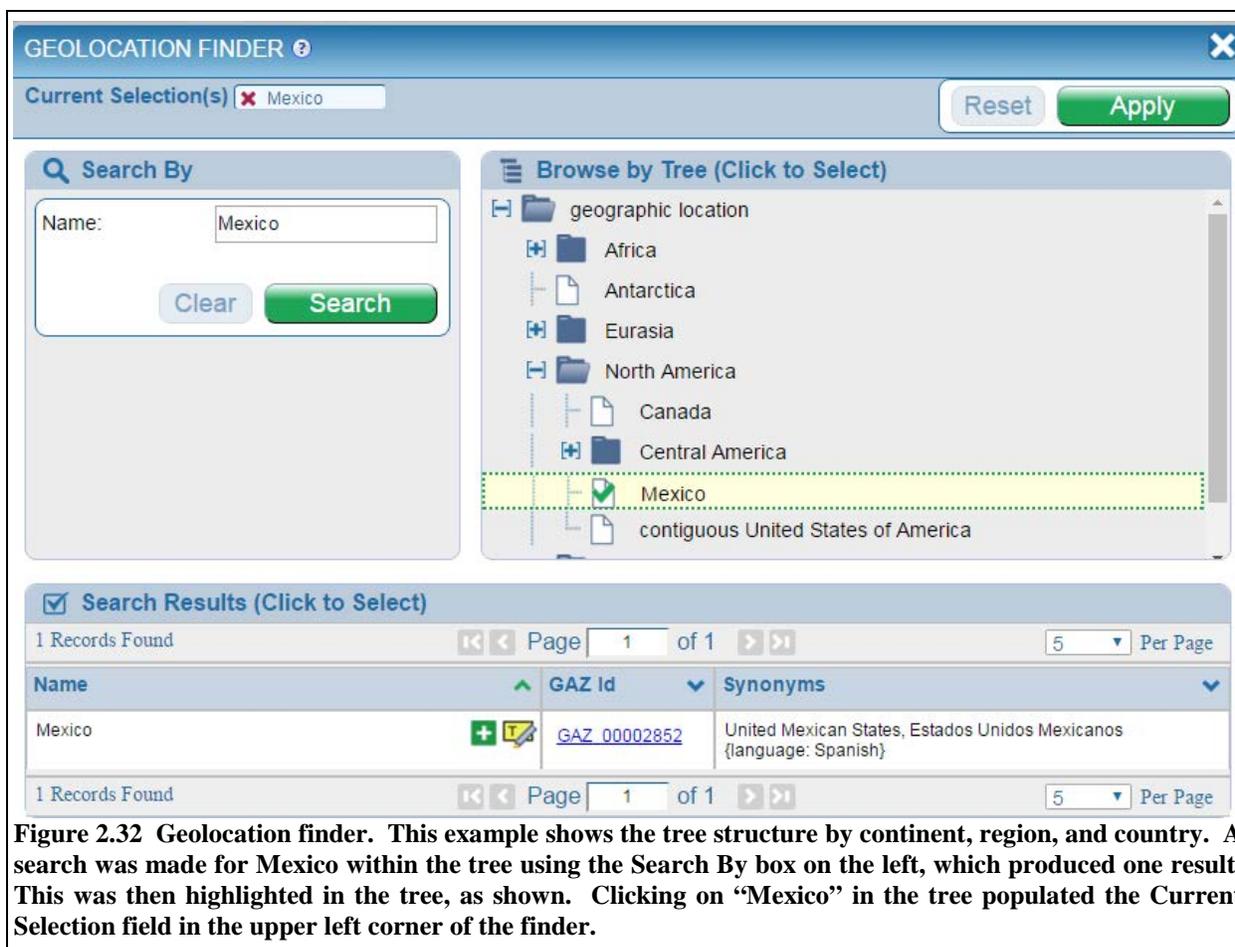
**Figure 2.31 Organism Finder showing a search for dengue with dengue highlighted in the tree**

### 2.2.4.6 Geolocation Finder

This feature combines the locations where study subjects were studied and the external ontology known as Gazetteer (<http://bioportal.bioontology.org/ontologies/1397>). It can be found on the B Cell Search, T Cell Search, and MHC Ligand Specialized search pages under the Host Details subsection within the Immunization section. Gazetteer represents geographic locations in a formal ontological hierarchy. The IEDB's Geographic Location Finder was created by taking the geographic locations mentioned in the manuscripts containing data curated by the IEDB and presenting them in the formal ontological hierarchy of Gazetteer. Thus, every geographic location in the Finder links out to a GAZ ID, where additional information provided by Gazetteer such as definitions and synonyms can be found.

To use the Finder, simply navigate the tree by clicking open the continent of choice to find the country(ies) where study subjects were exposed to the immunogen tested in the immune epitope assays. Alternatively, one may use the "Search By" feature and type free text to search the tree. The Search By feature accepts synonyms as search criteria. Its use is shown in Figure 2.32 for the term "Mexico".

It must be noted that curators only capture geographic location information when authors explicitly state that subjects were recruited in a specific country. The location of authors' affiliations is not used to populate this field.



**Figure 2.32 Geolocation finder.** This example shows the tree structure by continent, region, and country. A search was made for Mexico within the tree using the Search By box on the left, which produced one result. This was then highlighted in the tree, as shown. Clicking on “Mexico” in the tree populated the Current Selection field in the upper left corner of the finder.

## 2.3 Tools

The Analysis Resource pull-down menu at [www.iedb.org](http://www.iedb.org) accesses the IEDB Analysis Resource and has five items – Analysis Resource Overview, T Cell Epitope Prediction, B Cell Epitope Prediction, Epitope Analysis Tools, and Tool Licensing Information. The purpose of the Analysis Resource of the IEDB is to provide computational tools that enhance the value of the IEDB database to the user. Providing access to tools in one centralized location helps make users aware of available solutions to their problems. All of the information contained within the Analysis Resource, including analysis tools and algorithms developed by the IEDB staff, are freely available to the scientific community. For commercial entities that wish to access the tools in-house, licensing information for the tools can be accessed from the pull-down menu. The Analysis Resource can also be accessed via the rightmost panel on the IEDB home page, although only the most frequently used tools are listed.

The tools provided in the analysis resource fall into two categories – prediction tools and analysis tools. Prediction tools extrapolate beyond data held in the database. They can be used to predict epitopes in protein sequences or predict properties of known epitopes, such as their MHC binding affinity. Analysis tools help extract and interpret data contained in the database.

For prediction tools, it is important to differentiate between the **tool** making predictions, and the **method** used to generate that tool, given a set of training data. For example, the artificial neural network method or approach, when trained on a particular data set, will yield a prediction tool. As the ANN method is trained on different data sets, different corresponding prediction tools will result. In this way, the ANN

method can be used to develop an MHC class I prediction tool and a separate MHC class II prediction tool. These tools can be refined as more data are available for training. One benefit of the IEDB is that it allows implementing methods to automatically generate new prediction tools as the database grows.

Prediction tools can be subdivided into categories by what they predict. The current tools fall into the subcategories listed below:

- T cell epitopes – MHC class I and II binding prediction
- T cell epitopes – Processing prediction
- T cell epitopes – Immunogenicity prediction
- B-cell epitope prediction

The AR Labs page (<http://tools.iedb.org/main/labs/>) was introduced in early 2016. Tools under AR Labs are experimental and are considered to be beta versions, not quite ready for production yet. They are intended for further research, updates and testing. This means they are functional, but the team is still actively developing on improving them and more substantial changes are expected in upcoming releases.

The next subsections will describe each prediction and analytical tool in more detail.

## **2.3.1 T Cell Epitope Prediction**

### **2.3.1.1 T Cell Epitopes - MHC binding prediction**

The Analysis Resource provides tools for predicting peptide binding to MHC class I and II molecules. For class I binding predictions, users can select predictions performed with tools derived from ten different methods – IEDB Recommended, artificial neural network (ANN), stabilized matrix method (SMM), SMM with a peptide:MHC binding energy covariance matrix (SMMPMBEC), Scoring matrices derived from combinatorial peptide libraries (Comblib\_Sidney2008), Consensus, NetMHCpan, PickPocket, NetMHCcons, and NetMHCstabpan (introduced in 2016). The Consensus method is derived from the ANN, SMM, and CombLib methods. For class II binding predictions, users can select IEDB Recommended, the Combinatorial Library method, the NN\_align, the SMM\_align method, a method devised by Sturniolo et al. and used in TEPITOPE, NetMHCIIpan, and a consensus method derived from the NN-align, SMM\_align, Sturniolo methods, and the Combinatorial Library. Tutorials and example data are available for both the class I and II tools.

A RESTful interface is available for the MHC class I and class II prediction tools. This allows users to perform predictions on the IEDB server in batch mode without having to install any software on their own systems. Additionally, users will always be assured that they are using the latest version of the tools. More information is available in Section 2.3.5. These tools are also available in standalone form on the Download tab and can be downloaded at <http://tools.iedb.org/main/download/>.

#### **2.3.1.1.1 Peptide Binding to MHC Class I Molecules**

Users can select from ten different methods for predicting class I epitopes – ANN, SMM, SMMPMBEC, Comblib\_Sidney2008, Consensus, NetMHCpan, NetMHCcons, PickPocket, NetMHCstabpan, and IEDB recommended, which are described further below. A check box can be used to show only frequently occurring alleles. This allows the selection of only those alleles that occur in at least 1% of the human population or allele frequency of 1% or higher. However, un-checking the check-box will allow selection of all the alleles and corresponding peptide lengths for a particular species. Users can also upload an allele file instead of entering allele on the page one at a time.

### **Artificial Neural Network**

Artificial neural networks (ANN) are computer algorithms modeled after the brain. They consist of many simple processing units which are wired together in a communication network. Each unit is a simplified model of a neuron which sends off a new signal if it receives a sufficiently strong input signal from the other units to which it is connected. The strength of these connections can be varied in order for the network to perform a desired pattern of node signal activity, which is learned from a set of input training data.

ANNs have been trained for 81 different Human MHC alleles including HLA-A, -B, -C and -E. Furthermore, predictions for 41 animal (Monkey, Cattle, Pig, and Mouse) alleles are available. The current version of ANN is based on NetMHC 4.0. The method is describe in Nielsen et al. (Protein Science, 2003) and Andreatta and Nielsen (Bioinformatics, 2016).

### **Stabilized Matrix Method (SMM)**

The Stabilized Matrix Method (SMM) described by Peters and Sette (BMC Bioinformatics, 2005) can be applied to calculate matrices from quantitative affinity data of peptides binding to MHC molecules. The advantage of this method is that it suppresses the noise present in the training data, caused by the inevitable experimental error as well as the limited number of data points.

### **Stabilized matrix Method with a Peptide:MHC Binding Energy Covariance matrix (SMMPMBEC)**

SMMPMBEC is an improved version of SMM. It is different from SMM in that it addresses sparseness of peptide sequence coverage that is often found in binding data sets by using Peptide:MHC Binding Energy Covariance (PMBEC). The PMBEC matrix was derived from experimentally determined binding affinity measurements using combinatorial peptide libraries. SMMPMBEC is described in Kim et al. BMC Bioinformatics 2009.

### **Scoring matrices derived from combinatorial peptide libraries (Comblib\_Sidney2008)**

Comblib\_Sidney2008 refers to a set of predictors (i.e. scoring matrices) that were derived from binding affinity measurements of combinatorial peptide libraries against a panel of MHC alleles. This work is described in Sidney et al. (Immunome Res. 2008). This class of predictors is unique in that average binding energy contribution of a given residue at a position is directly measured, without worrying about limited peptide sequence coverage.

### **Consensus**

The Consensus predictor was motivated by an idea that predictions made by consulting “consensus” of individual predictions from multiple predictors may result in improved performance over that of any individual ones. For MHC-I, a work describing an early implementation can be found in Moutaftsi M et al. (Nat Biotech 2006). The methods used for Consensus are ANN, SMM, and CombLib\_Sidney2008. The Consensus method uses as many of these three component methods as possible, depending on their applicability for the chosen allele and length.

### **NetMHCpan**

NetMHCpan version 3.0 predicts binding of peptides to any MHC class I molecule of known sequence using artificial neural networks (ANNs). The method is trained on more than 180,000 quantitative binding data covering 172 MHC molecules from human (HLA-A, B, C, E), mouse (H-2), cattle (BoLA), primates (Patr, Mamu, Gogo) and swine (SLA). It has been trained on an extended data set of 8 – 13mer peptides using a method described by Andreatta and Nielsen (Bioinformatics, 2016). Further information about the method can be found in Hoof I et al. (Immunogenetics 2009) and Nielsen and Andreatta (Genome Medicine, 2016).

### **PickPocket**

PickPocket predicts binding of peptides to any known MHC molecule using position specific weight matrices. The method is trained on more than 150,000 quantitative binding data covering more than 150 different MHC molecules. Predictions can be made for HLA-A, B, C, E and G alleles, as well as for non-human primates, mouse, cattle, and pig. Version 1.1 has been retrained on extended data set including 10 prevalent HLA-C and 7 prevalent BoLA MHC-I molecules. The matrices of pocket-library are generated using the SMM\_pmbec method described in Kim et al. (BMC Bioinformatics, 2009). Predictions can be made for 8 – 12mer peptides. All non-9mer predictions are made using approximations. The prediction values are given in nM IC50 values. The algorithm is described in Zhang et al. (Bioinformatics, 2009).

### **NetMHCcons**

NetMHCcons predicts binding of peptides to any known MHC class I molecule. This is a consensus method for MHC class I predictions integrating three state-of-the-art methods, NetMHC, NetMHCpan, and PickPocket, to give the most accurate predictions. NetMHC is an artificial neural network-based (ANN) allele-specific method which has been trained using 94 MHC class I alleles. Version 3.4 is used as part of NetMHCcons-1.1. NetMHCpan is a pan-specific ANN method trained on more than 115,000 quantitative binding data covering more than 120 different MHC molecules. Version 2.8 is used as part of NetMHCcons-1.1. PickPocket Version 1.1 is described above. NetMHCcons can generate predictions for peptides of 8 – 15 residues in length. The prediction values are given in nM IC50 values. The method is described in detail in Karosiene et al. (Immunogenetics, 2012).

### **NetMHCstabpan**

NetMHCstabpan was introduced in 2016. It predicts binding stability of peptides to any known MHC molecule using artificial neural networks (ANNs). The method is trained on more than 25,000 quantitative stability data covering 75 different HLA molecules. The user can upload full length MHC protein sequences, and have the server predict MHC restricted peptides from any given protein of interest. Predictions can be made for 8-14 mer peptides, although all non 9mer predictions are made using approximations. Most HLA molecules have a strong preference for binding 9mers. The prediction values are given in half life time in hours values and as %-Rank to a set of 200,000 random natural peptides. The method is described in Rasmussen et al. (Journal of Immunology, 2016).

### **IEDB Recommended**

IEDB recommended is the default prediction method selection. It considers all alleles and their corresponding peptide lengths for a particular species. For each allele-length combination, consensus method is used, which includes ANN, SMM, and CombLib. If none of these methods are available for the allele, NetMHCpan is used.

#### **2.3.1.1.2 Peptide Binding to MHC Class II Molecules**

Users can select from seven different methods for predicting class II epitopes – IEDB recommended, SMM-align, Sturniolo, Combinatorial Library, Consensus, NN-align, and NetMHCIIpan. By default, the overall best method (IEDB recommended) is selected. However, not all methods can currently make predictions for all alleles, so only the alleles available will be displayed. The seven methods are described further below.

### **SMM-align**

The MHC class II binding groove is open at both ends making the correct alignment of a peptide in the binding groove a crucial part of identifying the core of an MHC class II binding motif. The stabilization matrix alignment method, SMM-align, allows for direct prediction of peptide:MHC binding affinities. The method uses amino terminal peptide flanking residues (PFR) to get a consistent gain in predictive performance by favoring binding registers with a minimum PFR length of two amino acids. The method predicts quantitative peptide:MHC binding affinity values. The method has been trained and evaluated on

a data set that covers the nine HLA-DR supertypes suggested and three mouse H2-IA allele. The method is described by Nielsen et al. (BMC Bioinformatics, 2007).

### **Sturniolo**

This matrix-based approach is used in the TEPITOPE class II epitope prediction program. It is described in Sturniolo et al. (Nat. Biotechnol., 1999).

### **Combinatorial Library**

The positional scanning combinatorial libraries approach utilized a pool of random peptide libraries to systematically measure the contribution to MHC binding from each amino acid at each of the nine positions at the binding peptide. Each pool in the library contains 9-mer peptides with one fixed residue at a single position. With each of the 20 naturally occurring residues represented at each position along the 9-mer backbone, the entire library consisted of 180 peptide mixtures. Competitive binding assays were then carried out to determine the IC50 values for each pool. IC50 values for each mixture were standardized as a ratio to the geometric mean IC50 value of the entire set of 180 mixtures, and then normalized at each position so that the value associated with the optimal value at each position corresponds to 1. For each position, an average (geometric) relative binding affinity (ARB) was calculated, and then the ratio of the ARB for the entire library to the ARB for each position was derived. The final results are a set of 9 by 20 scoring matrices which could predict the binding of novel peptides to MHC molecules.

The methodology is the same as that used for the MHC class I combinatorial library tool as described in Sidney et al. Immunome Res. 2008. A description of the ARB method is contained in Bui et al., Immunogenetics, 2005.

### **Consensus**

The consensus method was developed by the IEDB team by exploiting features of three aforementioned methods. The Consensus method uses NN-align, SMM-align, and the combinatorial peptide scanning library. When the scanning library is not available for an allele, the Sturniolo method is used instead. A paper describing the original method was published by Wang et al. (PLoS Comput Biol, 2008), and a subsequent paper was published by Wang et al. (BMC Bioinformatics, 2010). The datasets used in assessing the performance of the NN-align, SMM-align, and Sturniolo methods and in developing the Consensus method, as described in Wang et al. (2010), can be found at <http://tools.iedb.org/main/datasets/>.

### **NN-align**

NN-align is an artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. It simultaneously identifies the MHC class II binding core and binding affinity. The method is trained using an algorithm that corrects bias in the training data caused by redundancy in binding core representation. Prediction accuracy has been shown to improve significantly when information about the residues flanking the peptide-binding core is taken into account. A 2009 paper in BMC Bioinformatics by Nielsen and Lund describe the method in detail (PMID: 19765293).

### **NetMHCIIpan**

NetMHCIIpan predicts binding of peptides for all three human MHC class II isotypes- HLA-DR, HLA-DP, and HLA-DQ, in addition to mouse H-2-I molecules. A paper describing the method was published by Nielsen et al. in PLoS Computational Biology, 2008, and a paper describing the latest version 3.1 was published by Andreatta et al. in Immunogenetics, 2015.

### **IEDB Recommended**

IEDB recommended is the default prediction method selection. It considers all alleles and their corresponding peptide lengths for a particular species. For each allele-length combination, consensus method is used, as described above. If none of these methods are available for the allele, NetMHCIIpan is used.

### **2.3.1.2 T Cell Epitopes – MHC I Processing Prediction**

#### ***2.3.1.2.1 Proteasomal cleavage/TAP transport/MHC class I combined predictor***

For the prediction of antigen processing through the MHC class I antigen presentation pathway, we incorporated predictions of proteasomal cleavage and TAP transport similar to the MHCPathway website described in (Tenzer et al, CMLS, 2005). The predictions are based on in vitro experiments characterizing the sequence specificity of proteasomal cleavage and TAP transport. The goal of the prediction is to identify MHC-I ligands (peptides that are naturally processed from their source proteins and presented by MHC class I molecules).

The proteasomal cleavage predictions evaluate how efficiently a peptide or its N-terminally prolonged precursors can be liberated from its source protein. The TAP transport predictions evaluate how efficiently a peptide or its N-terminal prolonged precursors are transported into the ER by TAP (Peters et al., Immunol, 2003). When this information is taken together and combined with MHC class I binding predictions, the tool yields a prediction of the efficiency with which a peptide is presented on the cell surface. The methods available are ANN, SMM, SMMPMBEC, Comblib\_Sidney2008, NetMHCpan, PickPocket, NetMHCcons, and IEDB Recommended, which are described in Section 2.3.1.1.1.

A check box can be used to show only frequently occurring alleles. This allows the selection of only those alleles that occur in at least 1% of the human population or allele frequency of 1% or higher. However, un-checking the check-box will allow selection of all the alleles and corresponding peptide lengths for a particular species. Users can also upload an allele file instead of entering allele on the page one at a time.

#### ***2.3.1.2.2 Neural network based prediction of proteasomal cleavage sites (NetChop) and T cell epitopes (NetCTL/NetCTLpan)***

NetChop produces neural network predictions for cleavage sites of the human proteasome (Kesmir et al., 2002). NetChop takes into account the characteristics of the structurally modified proteasomes found in cells stimulated by gamma-interferon under physiological conditions. The NetChop algorithm was trained on in vitro data and MHC Class I ligand data. The use of this training set, combined with the artificial neural network methodology, makes the prediction of cleavage sites more accurate. NetChop has been trained only on human data, but since the proteasome structure is quite conserved, the algorithm developers believe that the tool is capable of making reliable predictions for at least the other mammalian proteasomes.

NetCTL predicts CTL epitopes in protein sequences integrating prediction of peptide MHC binding, proteasomal C terminal cleavage and TAP transport efficiency. The method is described in detail in Larsen et al. (Eur J Immunol., 2005). NetCTLpan is an update to the original NetCTL server that allows for prediction of CTL epitope with restriction to any MHC molecules of known protein sequence (Stranzl et al., Immunogenetics, 2010). This tool is also available in a standalone form on the Download tab and can be downloaded at <http://tools.iedb.org/main/download/>.

### **2.3.1.2.3 MHC-NP – Prediction of peptides naturally processed by the MHC**

MHC-NP is a tool for predicting peptides naturally processed by the MHC pathway and was introduced to the Analysis Resource in 2013. MHC-NP employs data obtained from MHC elution experiments in order to assess the probability that a given peptide is naturally processed and binds to a given MHC molecule. This tool was the winner of the 2nd Machine Learning Competition in Immunology, where it yielded state-of-the-art accuracy for the prediction of peptides eluted from human HLA-A\*02:01, HLA-B\*07:02, HLA-B\*35:01, HLA-B\*44:03, HLA-B\*53:01, HLA-B\*57:01 and mouse H2-D(b) and H2-K(b) MHC molecules. The method is described in Giguère et al. (J Immunol Methods, 2013). Because the tool was developed by researchers external to the IEDB team, it is not included in the virtual machine image.

### **2.3.1.3 T cell class I pMHC immunogenicity predictor**

An MHC class I immunogenicity tool was introduced in 2013. T-cells have to recognize peptides presented on MHC molecules to be activated and elicit their effector functions. Several studies have demonstrated that some peptides are more immunogenic than others and therefore more likely to be T-cell epitopes. This tool is based on an analysis of a large set of data describing the immunogenicity of peptides presented on various MHC-I molecules.

Two main conclusions could be drawn from this analysis. First, in line with previous observations, the developers showed that positions P4-6 of a presented peptide are more important for immunogenicity. Second, some amino acids, especially those with large and aromatic side chains, are associated with immunogenicity. This information was combined into a simple model that was used to demonstrate that immunogenicity is, to a certain extent, predictable. This model was validated with data from two independent epitope discovery studies. This model indicated that T-cells are equipped to better recognize viral than human (self) peptides. The method is described in Calis et al. (PLoS Comput Biol., 2013). This tool is also available in a standalone form on the Download tab and can be downloaded at <http://tools.iedb.org/main/download/>.

## **2.3.2 B Cell Epitope Prediction**

### **2.3.2.1 Prediction of linear epitopes from protein sequence**

Six different tools are provided that predict antibody epitope candidates from amino acid sequences. Five are based on amino acid property scales and a sixth method uses a Hidden Markov Model. Parameters such as hydrophilicity, flexibility, accessibility, and antigenic propensity of polypeptides chains have been correlated with the location of continuous epitopes in a few well-characterized proteins. Based on these observations, amino acid property scales have been developed to predict antigenic determinants. Each scale consists of 20 values assigned to each of the amino acid residues on the basis of their relative propensity to possess the property described by the scale. The following amino acid property scales have been selected and implemented based on their popularity and coverage of different categories.

- Secondary structure - Chou and Fasman beta turn prediction
- Surface exposure - Emini surface accessibility prediction
- Flexibility - Karplus and Schulz flexibility prediction
- Antigenicity - Kolaskar and Tongaonkar antigenicity prediction
- Hydrophobicity/hydrophilicity - Parker hydrophilicity prediction

BepiPred combines the predictions of a hidden Markov model and the propensity scale of Parker et al. (Biochemistry, 1986). It is described in Larsen et al. (Immunome Research, 2006). The linear B cell

epitope predictor is also available as a standalone tool. Further information can be found on the Download tab at <http://tools.iedb.org/main/download/>.

### **2.3.2.2 DiscoTope - Prediction of epitopes from protein structure**

DiscoTope was augmented in 2013 to include include version 2.0 as well as version 1.1. DiscoTope is designed specifically to predict discontinuous epitopes. It uses protein three-dimensional structural data in addition sequence data. The method is based on amino acid statistics, spatial information, and surface accessibility in a compiled data set of discontinuous epitopes determined by X-ray crystallography of antibody/antigen protein complexes. DiscoTope 2.0 works with a novel definition of the spatial neighborhood used to sum propensity scores and half-sphere exposure as a surface measure. The results are more accurate than version 1.1 but involve more complex calculations that increase run times by a factor of approximately 30.

DiscoTope 1.1 is described in Haste Andersen et al. (Protein Sci., 2006), and DiscoTope 2.0 is described in Kringelum et al. (PLoS Comp. Bio, 2012).

### **2.3.2.3 ElliPro - Epitope prediction based upon structural protrusion**

ElliPro predicts linear and discontinuous antibody epitopes based on a protein antigen's 3D structure. ElliPro accepts either a protein structure (preferred) or a protein sequence as an input. If a protein sequence is used, ElliPro will predict its 3D structure by homology modeling. Its use is described in the Tutorial tab of the ElliPro section of the Analysis Resource. The method is described in Julia Ponomarenko et al. (BMC Bioinformatics, 2008). It is also available as a standalone tool. Further information can be found on the Download tab at <http://tools.immuneepitope.org/ellipro/download/>.

### **2.3.2.4 Prediction of ImmunoGlobulin Structure (PIGS)**

PIGS is a web server for the automatic modeling of immunoglobulin variable domains based on the canonical structure method. It takes heavy and light chain sequences as input, automatically selects the best template, and provides a complete three-dimensional model of the target antibody as an output that can be downloaded or displayed on-line. PIGS was developed by Paolo Marcatili and Anna Tramontano at the Biocomputing Unit of Sapienza University in Rome. A detailed description of PIGS can be found in Marcatili et al. (Bioinformatics, 2008).

### **2.3.2.5 Methods for modeling and docking of antibody and protein 3D structures**

This web page provides information on methods for modeling and docking of antibody and protein 3D structures. The methods are available as web servers and are hosted outside the IEDB except one method, PIGS, described above. One of the approaches at B-cell epitope prediction involves antibody-antigen docking, or 3D modeling of an antibody-antigen complex, for known 3D structures of both antibody and antigen. When the structures of either antibody or protein or both are unknown (that is, not available in PDB) but sequences are known, the user can attempt to obtain the 3D structural models of the antibody and the antigen, following the chart displayed on the web page and using the described web servers. Links are provided to these servers for the user's convenience.

## **2.3.3 Epitope Analysis Tools**

### **2.3.3.1 Population coverage**

T cells recognize a complex between a specific MHC type and a particular pathogen-derived epitope and thus a given epitope will elicit a response only in individuals that express an MHC molecule capable of binding that particular epitope. MHC molecules are extremely polymorphic (over a thousand different variants are known in humans). Therefore, selecting multiple peptides with different MHC binding specificities will afford increased coverage of the patient population targeted as vaccine recipients. The

issue of population coverage in relation to MHC polymorphism is further complicated by the fact that different MHC types are expressed at dramatically different frequencies in different ethnicities. Thus, without careful consideration, a vaccine with ethnically biased population coverage could result. To address this issue, the actual/predicted binding capacity of potential epitopes to as many different MHC molecules possible (and when available, also restriction data of T cell responses recognizing the epitope) can be used to project the population coverage in different ethnicities of different vaccine candidates or epitope sets. Accordingly, epitope-based vaccines or diagnostics can be designed to maximize population coverage, while minimizing complexity (that is, the number of different epitopes included in the diagnostic or vaccine), and also minimizing the variability of coverage obtained or projected in different ethnic groups.

An important consideration in the process of epitope selection is that the patient population coverage afforded by a given set is not simply corresponding to the sum of the coverage of its individual components. Thus, to calculate the coverage afforded by a given mixture of epitopes, a more comprehensive approach and a suitable algorithm has been developed for this specific purpose (Bui et al. BMC Bioinformatics 2006). This method calculates the fraction of individuals predicted to respond to a given epitope set on the basis of HLA genotypic frequencies, assuming non-linkage disequilibrium between HLA loci, and on the basis of MHC binding and/or T cell restriction data. The algorithm is briefly explained here. First, genotypic frequencies of various MHC are tabulated. Each time a peptide binds to a given MHC, a “hit” is recorded for that MHC. The process is repeated for all peptides. Then the hits for MHC are tallied. Next, the frequency of each possible diploid MHC combination (phenotype) is calculated. For  $n$  MHC types, this corresponds to an  $n \times n$  tabulation of the frequency at which each specific pair of MHCs will be found in the population from which the MHC frequencies are derived. A similar table is generated to contain the number of hits per each of the MHC combinations by adding the number of hits associated with each of the two alleles of MHC in the combination (a simple exception is the case of homozygous combinations, where the number of hits is simply the number of hits of the given MHC). From these two tables, a frequency distribution is assembled, tabulating the genotypic frequency of all MHC combinations associated with a certain number of hits. The result of the analysis is displayed as a frequency distribution histogram and a cumulative frequency plot.

HLA allele genotypic frequencies were obtained from Allele Frequency Net Database (<http://www.allelefreqencies.net/>). At present, Allele Frequency Net Database (Gonzalez-Galarza FF et al., Nucleic Acids Res. 2015) provides allele frequencies for 115 countries and 20 different ethnicities grouped into 16 different geographical areas. In addition, the program also accepts custom populations with allele frequencies defined by users. Multiple population coverages can be simultaneously calculated and an average population coverage is generated. Since MHC class I and MHC class II restricted T cell epitopes elicit immune responses from two different T cell populations (CTL and HTL, respectively), the program provides three calculation options to accommodate different coverage modes: (1) class I separate, (2) class II separate, and (3) class I and class II combined. For each population coverage, the tool computes the following: (1) projected population coverage, (2) average number of epitope hits / HLA combinations recognized by the population, and (3) minimum number of epitope hits / HLA combinations recognized by 90% of the population (PC90).

### **2.3.3.2 Epitope conservancy**

In a diagnostic or epitope-based vaccine setting, focusing on conserved epitopes allows for targeting responses around pathogen variability, whether it exists prior to infection, or develops in the natural course of disease. The use of conserved epitopes would be expected to focus the immune response on sequences crucial for retaining biological function of the pathogen proteins, and thus with intrinsically lower variability, even under immune pressure. The epitope conservancy analysis tools implemented here aims to address the issue of variability (or conservation) of epitopes, and to assist in the selection of epitopes with the desired pattern of conservation. The algorithm has been implemented to calculate the

degree of conservancy of an epitope within a given protein sequence set at different degree of identities. The degree of conservation is defined as the number of protein sequences that contain the epitope at a given identity level, divided by the total number of protein sequences found in the dataset analyzed (Bui et al. BMC Bioinformatics 2007).

A new feature was added in the past year. User can now indicate whether they want duplicated protein sequences removed from the results set by checking a box in the “Specify calculation options” section of the input page. Previously redundant results were removed. For users wanting to study different populations, keeping redundant sequences can be very informative.

### 2.3.3.3 Epitope Cluster Analysis

The Epitope Cluster Analysis tool (Kim et al., Nucleic Acids Res. 2012) groups epitopes into clusters based on sequence identity. A cluster is defined as a group of sequences that has a sequence similarity greater than the minimum sequence identity threshold specified. Epitope sequences can be either directly entered in the text area or uploaded from a file. Two acceptable sequence formats are PLAIN and FASTA. The user can select the sequence identity threshold at which they want to calculate epitope clusters. Clusters are displayed in a table format where clusters are indicated by table rows which have the same color. All calculated cluster results can be saved to a file by clicking on the "Download data to file" button.

### 2.3.3.4 Computational Methods for Mapping Mimotopes to Protein Antigens

A web page devised by Dr. Julia Ponomarenko was added to the Analysis Tools section in 2014 that provides guidelines for mapping mimotopes to protein antigens. Mimotopes are molecules, often peptides, which are recognized by antibodies or T cells that were primed to recognize a different epitope. For antibody epitope mapping, peptide mimotopes can be selected from a combinatorial library (usually peptides from a specific protein antigen) on binding to a specific monoclonal antibody with high affinity. They are commonly obtained using phage display. Experimental data on mimotopes can be found in the IEDB, MimoDB (<http://immunet.cn/mimodb/index.html>), and journal publications. Mimotopes can mimic essential features of genuine epitopes and therefore their mapping to the antigen is a logical and essential step of the mimotope analysis. This web page provides information on available methods for mimotope mapping, how to search the IEDB for mimotopes, and an example of a mimotope dataset and the results of its mapping, using the available web servers hosted outside the IEDB.

Methods available as web servers and operational as of July 2014 (they all require the user submitting an antigen protein 3D structure, providing either PDB ID of the structure or the pdb file) are as follows:

- **MIMOX** (<http://immunet.cn/mimox/>)
- **Pepitope server** (<http://pepitope.tau.ac.il/>) provides three algorithms: Mapitope, PepSurf, or combined
- **EpiSearch** (<http://curie.utmb.edu/episearch.html>)

The web page also lists other published methods that are unavailable as web servers. They include the following:

- **PepMapper** (<http://www.ncbi.nlm.nih.gov/pubmed/22701536>) (2012) - provided URL didn't respond
- **MimoPro** (<http://www.ncbi.nlm.nih.gov/pubmed/21609501>) (2011) - provided URL didn't respond
- **Pep-3D-Search** (<http://www.ncbi.nlm.nih.gov/pubmed/19087303>) (2008) - available for download at <http://kyc.nenu.edu.cn/Pep3DSearch/>

- **MEPS** (<http://www.ncbi.nlm.nih.gov/pubmed/17430573>) (2007)
- **MIMOP** (<http://www.ncbi.nlm.nih.gov/pubmed/16434442>) (2006)
- **3D-Epitope-Explorer (3DEX)** (<http://www.ncbi.nlm.nih.gov/pubmed/15834923>) (2005) - available for download at <http://www.schreiber-abc.com/3dex/>
- **Enshell-Seijffers et al.** (<http://www.ncbi.nlm.nih.gov/pubmed/14596802>) (2003)
- **FINDMAP** (<http://www.ncbi.nlm.nih.gov/pubmed/12935344>) (2003)
- **SiteLight** (<http://www.ncbi.nlm.nih.gov/pubmed/12824481>) (2003)

The webpage also contains instructions on how to search for mimotopes in the IEDB.

### 2.3.4 IEDB Analysis Resource Labs

The AR Labs page (<http://tools.iedb.org/main/labs/>) was introduced in early 2016. Tools under AR Labs are experimental and are considered to be beta versions, not quite ready for production yet. They are intended for further research, updates and testing. This means they are functional, but the team is still actively developing on improving them and more substantial changes are expected in upcoming releases. Therefore users are asked not to expect the same level of stability for these tools as for other tools in the IEDB analysis resource. Users are welcome to send bug reports and suggestions for improvements for these tools via <http://help.iedb.org/hc/en-us/requests..>

Two tools are currently available on the Labs web page:

**RATE (Restrictor Analysis Tool for Epitopes)**: is an automated method that can infer HLA restriction for a set of given epitopes from large datasets of T cell responses in HLA typed subjects. The tool takes two data files, one containing the alleles expressed by the subjects and the other containing the response of the peptides in the subjects. The tool calculates the odds ratio and estimates its significance using Fisher's exact test. It also calculates a parameter called relative frequency similar to odds ratio. The tool was developed with a focus on class II alleles but can also be applied to class I alleles. A description of the tool can be found in Paul et al. (Journal of Immunology, 2015)

**TepiTool**: The TepiTool provides prediction of peptides binding to MHC class I and class II molecules. The tool is designed as a wizard with six steps. Each field (except sequences and alleles) is filled with default recommended settings for prediction and selection of optimum peptides. The input parameters can be adjusted as per the user's specific needs. Users can go back to previous steps to change your selection before submission of the job. Once the user submits the job (at the end of step-6), they will not be able to make any more changes and will have to start the prediction all over again with updated input parameters. A description of the tool can be found in Paul et al. (Current Protocols in Immunology, 2016).

### 2.3.5 Benchmark references and data sets

The "Datasets" tab contains data sets and references related to benchmarking results of the IEDB's epitope prediction tools. There are three accordion sliders that can be expanded to reveal details. The three are for MHC class I binding prediction, MHC class II binding prediction, and B cell epitope prediction. In each case, the journal article title, authors, journal name, and year are listed along with a description of the data set, the date of its generation, details on the data set generation, the data format, and where the data set is available.

For the MHC class I binding predictions, the references are:

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

1;31(13):2174-81. doi: 10.1093/bioinformatics/btv123. Epub 2015 Feb 25. PubMed PMID: 25717196; PubMed Central PMCID: PMC4481849

- Kim Y, Sidney J, Buus S, Sette A, Nielsen M, Peters B. Dataset size and composition impact the reliability of performance benchmarks for peptide-MHC binding predictions. *BMC Bioinformatics*. 2014 Jul 14;15:241. doi: 10.1186/1471-2105-15-241. PubMed PMID: 25017736; PubMed Central PMCID: PMC4111843
- Kim Y, Sidney J, Pinilla C, Sette A, Peters B. Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. *BMC Bioinformatics*. 2009 Nov 30;10:394. PubMed PMID: 19948066.
- Peters B, Bui HH, Frankild S, Nielson M, Lundegaard C, Kostem E, Basch D, Lamberth K, Harndahl M, Fleri W, Wilson SS, Sidney J, Lund O, Buus S, Sette A., A community resource benchmarking predictions of peptide binding to MHC-I molecules, *PLoS Comput Biol*. 2006 Jun 9;2(6):e65. Epub 2006 Jun 9. PMID: 16789818

For the MHC class II binding predictions, the references are:

- Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*. 2010 Nov 22;11:568. PubMed PMID: 21092157; PubMed Central PMCID: PMC2998531.
- Wang, P., J. Sidney, C. Dow, B. Mothe, A. Sette, B. Peters. (2008). "A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach." *PLoS Computational Biology* 4(4). PMID: 18389056

For the B cell predictions, the references are:

- Kringelum JV, Nielsen M, Padkjær SB, Lund O. Structural analysis of B-cell epitopes in antibody:protein complexes. *Mol Immunol*. 2013 Jan;53(1-2):24-34. doi: 10.1016/j.molimm.2012.06.001. Epub 2012 Jul 10. PubMed PMID: 22784991; PubMed Central PMCID: PMC3461403
- Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol*. 2012;8(12):e1002829. doi: 10.1371/journal.pcbi.1002829. Epub 2012 Dec 27. PubMed PMID: 23300419; PubMed Central PMCID: PMC3531324.
- Rubinstein ND, Mayrose I, Martz E, Pupko T. Epitopia: a web-server for predicting B-cell epitopes. *BMC Bioinformatics*. 2009 Sep 14;10:287. doi: 10.1186/1471-2105-10-287. PubMed PMID: 19751513; PubMed Central PMCID: PMC2751785.
- Liang S, Zheng D, Standley DM, Yao B, Zacharias M, Zhang C. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics*. 2010 Jul 16;11:381. doi: 10.1186/1471-2105-11-381. PubMed PMID: 20637083; PubMed Central PMCID: PMC2910724.
- Ponomarenko JV, Bourne PE., Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *BMC Struct Biol*. 2007 Oct 2;7(1):64. PMID: 17910770

### 2.3.6 Other Ways to Access Tools

There are several ways that users can access many of the tools in the Analysis Resource in addition to using the tools directly on the Analysis Resource website. On the Analysis Resource home page (<http://tools.iedb.org>), there are tabs for Tools-API and Download. The former takes the user to

information about the RESTful interface. The RESTful (Representational State Transfer) web service is available for the MHC class I and II binding predictions, MHC class I processing predictions, and MHC-NP for T cell epitopes, and linear epitope prediction methods for B cells. The service sends POST request to the tools server, and relies on user supplied parameters. The Tools-API web page contains examples for accessing the service using the 'curl' program, although any program that sends POST requests to the server will work just as well (including a web browser). This feature allows users to perform predictions on the IEDB server in batch mode without having to install any software on their own systems. Additionally, users will always be assured that they are using the latest version of the tools.

The Download tab takes the user to the IEDB Tools Download web page. For users from non-commercial organizations with batch processing needs, the MHC class I and II binding prediction tools, linear B cell epitope prediction tool, ElliPro, the epitope cluster analysis tool, the MHC class I immunogenicity tool, and proteasomal cleavage prediction tool (added in 2016) are available as standalone scripts for download. These command line tools are kept in sync with the web tools and should therefore produce the same results as clicking through the web interface. They are freely available to non-profit/academic users through an open source license.

For commercial users that would like to run the analysis resource locally, licenses for a virtual machine image file or standalone downloads are available upon request for a modest annual fee. For the virtual machine, all tools are included except MHC-NP, and PIGS, which were not developed by the IEDB team. The virtual machine image is kept in sync with the current version of the IEDB Analysis Resource and is updated approximately every six months. Commercial licenses are also available. Requests for either license can be sent to [license@iedb.org](mailto:license@iedb.org).

## **2.4 Help Overview**

The Help pull-down menu contains items that aid the user to understand and utilize the features of the website. Users can access relevant documentation, request help, and provide feedback.

### **2.4.1 Support**

The IEDB Solutions Center is the primary resource for information on using the website's features. The user can submit help requests, check on the status of requests, browse and search the knowledge base and forums, and link to help documentation, such as the Curation Manual. In order to submit a help request via the Solutions Center, users must provide an email address, a subject line, and a description of the issue password. The Solutions Center can also be accessed by a link at the bottom of each web page.

The Solutions Center has extensive documentation organized in forums that address a variety of topics, including the epitope analysis and prediction tools, submitting data to the IEDB, querying the database, and general announcements and updates. The Solutions Center is also the repository for the Annual Compendia, the list of meta-analyses and other papers published by the IEDB team, release notes for the main website and the Analysis Resource, and the News and Noteworthy forum. In addition, written and video tutorials, including the presentations from the 2015 IEDB User Workshop can be found there.

### **2.4.2 Help Request**

Users can submit help requests in two ways. At the bottom of most pages is a "Help request" link. This will initiate a submission to the IEDB Solutions Center. The second method also utilizes the IEDB Solutions Center. The top menu bar on the IEDB Solutions Center web page has a "Help Request" link. Selecting this link brings the user to a "Submit a request" web page. The user must fill in their email address, the subject line, and a description of the help request. All help requests are forwarded to the same help desk.

Help requests should consist of problems that users have with the application, such as a certain function of the system not working. Questions on how to use features of the application should be covered in the online help available in the Solutions Center. After submitting a help request, a confirmation e-mail will be sent to the user's e-mail address, which will include the help request number. When the help request issue has been resolved, an e-mail will be sent to notify the user that the help request has been satisfied.

Help requests are generally responded to within one business day and solved within two business days. Requests that are actually bug reports by users are answered initially to inform the requestor that their issue is being forwarded to the appropriate developers and they will be notified when a fix is released. Requests involving the curation or re-curation of articles will typically take longer to resolve.

### **2.4.3 Provide Feedback**

All users are able to submit feedback, which will transmit an email message to the IEDB team. Feedback is intended for questions, input, and suggestions, such as new features they would like to see added in the future. Feedback helps the IEDB team update the system to provide users with the best possible experience. If users need help using the system or handling an unexpected result, a Help Request is more appropriate. The feedback feature can be accessed with the Support pull-down menu and the "Provide Feedback" at the bottom of each web page. Both initiate a submission to the IEDB Solutions Center.

### **2.4.4 Video Tutorials**

There is a link to the IEDB video tutorials under the Help menu at the top of each web page. This link takes the user to the "How-to Videos" article in the Solutions Center in the Tutorials and Reference Materials folder. Most of the videos are edited versions of live presentations made at the 2015 IEDB User Workshop held in October 2015 at the NIAID facility in Rockville, Maryland. The videos are listed in Table 2.1. An overview of the resource and information regarding the data structure and curation are presented first. The next category addresses the query and reporting features. Videos for all the tools in the Analysis Resource are listed next. Videos for the legacy website have been removed because the legacy site has been retired and can no longer be accessed by users.

Category	Topic
IEDB Overview, data curation, and data structure	Overview of the IEDB resource (19 minutes)
	IEDB data structure and curation (17 minutes)
Data query and reporting features	Finding Data Using the Home Page Search (29 minutes)
	Finding Data Using the Specialized Searches (12 minutes)
	The Immunome Browser (26 minutes)
	Home page search examples (20 minutes)
	Specialized search examples (20 minutes)
	Four infectious disease related queries (23 minutes)
	Two autoimmune disease related queries (15 minutes)
	Finding T and B cell epitopes in a given protein of interest (5 minutes)
	Example allergy query (3 minutes)
	IEDB and cancer epitopes (2 minutes)
	Queries and IEDB features relevant to vaccine development (2 minutes)
	Help and support features; database export options (25 minutes)
	IEDB Analysis Resource - Epitope Prediction and Analysis Tools, including epitope visualization
MHC class I binding prediction tools (41 minutes)	
MHC class I processing and immunogenicity tools (27 minutes)	
MHC class II binding prediction tools and API, including TepTool (62 minutes)	
TepiTool – a wizard interface for predicting MHC class I and II binding (20 minutes)	
B cell epitope prediction tools (67 minutes)	
Homology mapping tool and Epitope Viewer (23 minutes)	
B Cell Epitope Modeling and Docking, including Prediction of ImmunoGlobulin Structure (PIGS) (10 minutes)	
Analysis Tools (9 minutes)	
Tool downloads, epitope prediction references and benchmarks, and how to contribute tools to the IEDB (4 minutes)	

**Table 2.1 Available IEDB video tutorials**

## 2.5 More IEDB

The "More IEDB" pull-down menu contains links to other relevant information that may be of interest to users. Its component links are described below.

### 2.5.1 Database Export

The contents of the Immune Epitope Database are exported weekly to files in XML format. The database export page (Figure 2.32) contains a complete database export in XML format along with exports of the various records that are referenced in the complete database export. The complete dataset export is provided as a ZIP archive (iedb\_export.zip) containing one XML file for each reference contained in the database. The export also contains ZIP archives of XML files that provide users with the full record for the various identifiers that are referenced in the full database export, such as the IEDB source organism accession identifier list, the MHC allele name list, and the organism list. The supporting XML files only contain full records when the record is not otherwise available from the original source.

The database export page also provides the relevant XML Schema Definition (XSD) files for each of the XML files provided. The XSD file(s) for a particular XML file are located in the IEDB Schema box in the export table. In the case of the complete database export, multiple schemas are provided, with the primary schema being listed first, followed by any supporting schemas.

In addition to the XML export, the contents of the IEDB are exported weekly as a MySQL Structure Query Language (SQL) database. The database export page contains the complete database export in two formats, a single SQL script ([iedb\\_public.sql.gz](#)), and a TAR archive of MyISAM binary tables ([iedb\\_public.tar.gz](#)). Both export formats are generated from a [5.5.52-MariaDB](#) database server. The database export page also provides an entity-relationship diagram (ERD) for the MySQL database.

Archives of the data as they appeared in IEDB 1.0, 2.0, 2.1, and 2.3 are available by clicking on the *Archived Versions*. For each version, there is a compressed file that contains an XML for each reference. The corresponding XSD files are also available for download.

Select compilations of data are available for download from the Database Export web page in CSV format in the section titled CSV Metric Exports. Users can download these files to conveniently obtain all the information in the IEDB for specific categories. Files are available for all epitopes, antigens, T cell assays, B cell assays, MHC ligand assays, and references. In addition, an export for antibody and T cell receptors was added in 2016. In addition to the files available on the Database Export web page, comma separated values (CSV) file exports are available on-demand from all list pages and list tabs throughout the site.

Information about our database export features can be found in our [developer's forum](#).

XML Database Export	
<a href="#">Complete Database Export</a>	144MB
<a href="#">IedbAccessionList.zip</a>	37kB
<a href="#">MhcAlleleNameList.zip</a>	27kB
<a href="#">OrganismList.zip</a>	25MB
<a href="#">AssayTypeList.zip</a>	5kB
<a href="#">GeoLocList.zip</a>	2kB

IEDB Schema	
<a href="#">Curation.xsd (Primary IEDB schema)</a>	46kB
<a href="#">CurationSimpleTypes.xsd</a>	155kB
<a href="#">IedbPDBViewerSchema.xsd</a>	7kB
<a href="#">IedbAccessionList.xsd</a>	909B
<a href="#">MhcAlleleNameList.xsd</a>	1kB
<a href="#">OrganismList.xsd</a>	751B
<a href="#">AssayTypeList.xsd</a>	771B
<a href="#">GeoLocList.xsd</a>	642B

MySQL Database Export	
<a href="#">SQL Statement Export</a>	199MB
<a href="#">MyISAM Binary Export</a>	370MB

Physical Entity Relationship Diagram	
<a href="#">iedb_public_erd.pdf</a>	30kB

CSV Metric Exports	
<a href="#">epitope_full_v3.zip</a>	16MB
<a href="#">antigen_full_v3.zip</a>	804kB
<a href="#">tcell_full_v3.zip</a>	16MB
<a href="#">bcell_full_v3.zip</a>	20MB
<a href="#">mhc_ligand_full.zip</a>	21MB
<a href="#">reference_full_v3.zip</a>	10MB
<a href="#">receptor_full_v3.zip</a>	274kB

[Archived Versions](#)

**Figure 2.33 IEDB Database Export web page**

## 2.5.2 Meta-Analyses

The IEDB team has conducted a variety of analyses that examine data in the IEDB for several disease conditions and have been published in scientific journals. Over 20 papers have been published covering topics ranging from allergies, diabetes, and multiple sclerosis, to influenza, tuberculosis, malaria. When a user selects this menu item, they are directed to the Meta-analyses folder in the Knowledgebase and Forums section of the IEDB Solutions Center. The forum is updated as new meta-analyses are published. Users who subscribe to the folder will be automatically informed when content is added.

## 2.5.3 Citing the IEDB

Data and tools within the IEDB are presented as a public resource. Users are requested to consider citing

the IEDB when they present information obtained from the IEDB or use tools contained in the Analysis Resource. It is expected that the authors of an entry as well as the IEDB are properly cited whenever their work is referred to:

1. The IEDB website should be cited using the URL: <http://www.iedb.org>
2. The journal reference for the IEDB was updated after the release of IEDB 3.0. The IEDB should be cited as:

Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A, Peters B. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* 2014 Oct 9. pii: gku938. [Epub ahead of print] PubMed PMID: [25300482](https://pubmed.ncbi.nlm.nih.gov/25300482/).

## 2.5.4 Release Notes

Release notes are provided whenever a new release of the main website or the Analysis Resource is deployed to the public. When a user selects this menu item, they are directed to the Release Notes folder in the Knowledgebase and Forums section of the IEDB Solutions Center. The forum is updated as new releases are deployed. Users who subscribe to the folder will be automatically informed when content is added.

## 2.5.5 Links to External Sources

The IEDB system provides a list of links to external resources solely for the convenience of Immune Epitope Database visitors. The Immune Epitope Database has no interest in, responsibility for, or control over the linked-site. The Immune Epitope Database makes no promises or warranties of any kind, express or implied, including those of fitness for any particular purpose, as to the content of the linked-site. To view the links available, select the Links menu item. The hyperlinks on the links page are grouped by category:

- Public Databases, Prediction Algorithms, and Other Tools
- Antibody Related Links
- Bioinformatics Resource Centers
- MHC and TCR Related Links
- Protein Related Links
- Laboratory Resources
- Biodefense Resources

## 2.6 Learn More

The Learn More web page can be accessed from the link at the bottom of the Welcome box in the upper left-hand corner of the IEDB home page. The Learn More page has five major sections, as seen in Figure 2.33. The upper left section has a carousel of content from the New and Noteworthy forum in the Solutions Center. A graphic is usually included in the forum article that is displayed in the left-hand portion of the section. Below this section are three other sections titled Support, About the Data, and About Us. The Support section has links to submit a help request or feedback. The right-hand panel of the page lists the release notes for the most recent versions of the main website and the Analysis Resource.

**IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE**

Home | Specialized Searches | Analysis Resource | More IEDB

**News and Updates**

The redesigned IEDB 3.0 website was released as the production website (<http://www.iedb.org>) on 5 February 2015. It has been designed to be more user-friendly and more intuitive to use. A new query interface has been implemented for searches originating from the home page and from the specialized (advanced) search pages. A new version of the Immune Browser has also been implemented ...

**Support**

- Ask a question or submit an idea
- Browse support topics
- Watch how to videos
- Get help with common searches
- Browse links to epitope resources

**About the Data**

Export all or part of the IEDB  
Meta-analyses  
Read about the data fields  
See how the data is entered  
Learn about the IEDB ontology

**About Us**

- Citation guidelines
- IEDB publications
- Acknowledgments
- Terms of use
- Compendia & newsletters

**Latest Release Notes**

**IEDB 3.3 Release Notes (5 February 2015)**  
The redesigned IEDB website version 3.3.0 was deployed as the new production site on February 2015. Release highlights include:

...

**IEDB Analysis Resource v2.13 release notes**  
v2.13 - February 2015  
MHC I/II binding & processing

- (all) NetMHCcons 1.1 added
- (all) Pickpocket 1.1 added

...

**IEDB Analysis Resource v2.12 release notes**  
v2.12 - July 2014  
NEW Tool added!

- Prediction of Immunoglobulin Structures...

**IEDB v2.13.0 release notes**  
This topic contains the release notes for IEC version 2.13.0.

IEDB v2.13.0  
Enabled disease and allele finders...

[View Previous Release Notes >>](#)

[Provide Feedback](#) | [Help Request](#) | [Solutions Center](#)

Data Last Updated: March 01, 2015

**Figure 2.34** The Learn More page, accessed from the Welcome message in the upper left corner of the IEDB home page

## 2.6.1 Support

Each link is briefly described below.

### Ask a question or submit an idea

This link allows the user to submit a help request or suggestion via the IEDB Solutions Center. The user must specify their email address, the subject of their request, and description of the problem or suggestion. Users can also attach relevant files.

### Browse support topics

The IEDB Solutions Center home page has a variety of user help, including several resources for getting started with the IEDB, recent Release Notes, and timely articles.

### Watch how to videos

Videos describing the use of the main website and the Analysis Resource are located in the Solutions Center at <http://help.iedb.org/entries/140865-How-To-Videos>. See Section 2.4.4 for further details.

### Get help with common searches

There is a category in the Solutions Center that contains a list of common queries. These queries have been designed to parallel the IEDB meta-analysis efforts. Meta-analyses are performed at intervals to query the database for all data relating to subjects of interest (ex. Influenza, Malaria, Tuberculosis). These queries represent a simplified version of such analysis.

## **Browse links to epitope resources**

The IEDB system provides a list of links to external resources solely for the convenience of Immune Epitope Database visitors as described in Section 2.5.5.

## **2.6.2 About the Data**

Each link is briefly described below.

### **Export all or part of the IEDB**

Clicking on this link will bring the user to the Database Export page that is described in Section 2.5.1.

### **Meta-analyses**

The meta-analyses conducted by the IEDB team are described in Section 2.5.2.

### **Read about the data fields**

This link gives the user access to a subset of the Curation Manual hosted in a Wiki that describes and defines the various data fields. Using the Find function of their web browser, users can search for the data field name of interest in order to gain a fuller understanding of the field definition and the possible values.

### **See how the data is entered**

Users can access the full IEDB Curation Manual. The rules for curating all data fields are described in this Wiki.

### **Learn about the IEDB ontology**

Selecting IEDB Ontology takes the user to <http://ontology.iedb.org>, the home page for the Ontology of Immune Epitopes (ONTIE). The IEDB ontology is available to download in a format that can be viewed with the Protégé OWL editor. The file ONTIE.owl contains the classes and relations used in the IEDB export. IEDB-export.owl is the ONTIE-annotated export of the IEDB T cell epitope data. Smaller versions of the file are available for download. The ONTIE home page also lists related ontologies that are used in constructing ONTIE.

## **2.6.3 About Us**

### **Citation Guidelines**

This link brings the user to the “Citing the IEDB” web page described in Section 2.5.3.

### **IEDB Publications**

A list of publications relevant to the IEDB can be found on the Publications page. They have been grouped into four categories – General IEDB, Epitope Meta-analyses, Analysis Resource, and Curation. This list is updated with each new release of the main website. A list of IEDB related publications can also be found in the Publications folder in the IEDB Solutions Center at <http://help.iedb.org/hc/en-us/sections/114094004371-Publications>.

### **Acknowledgements**

A host of talented individuals have worked hard to make the Immune Epitope Database a reality. A roster of the current team members can be viewed on this page.

### **Terms of Use**

The Terms of Use page is a collection of statements that outline the conditions related to using the IEDB system. The Terms of use includes our privacy notice, copyright information, and various disclaimers.

The IEDB privacy policy describes what user information is collected, the circumstances for collecting it, and how it is used. Personal information about users is not collected unless the users choose to provide it.

### Annual Compendia

An archive of the IEDB Annual Compendia can be found at this link to the IEDB Solutions Center. The list includes all Annual Compendia going back to 2005. The section also includes the IEDB System and Architecture Design document.

## 3 Publications of the IEDB team by Year

Scientific articles written by the IEDB team members that are relevant to the IEDB are listed below by year of publication. In 2016, ten articles were published, 12 in 2015, seven in 2014, ten in 2013, six were published in 2012, three in 2011, nine in 2010, seven in 2009, nine in 2008, twelve in 2007, five in 2006, six in 2005, and one in 2004. In all, the team has written 97 scientific articles about the IEDB in the last 13 years.

### 3.1 2016

1. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, Clancy K, Courtot M, Derom D, Dumontier M, Fan L, Fostel J, Fragoso G, Gibson F, Gonzalez-Beltran A, Haendel MA, He Y, Heiskanen M, Hernandez-Boussard T, Jensen M, Lin Y, Lister AL, Lord P, Malone J, Manduchi E, McGee M, Morrison N, Overton JA, Parkinson H, Peters B, Rocca-Serra P, Ruttenberg A, Sansone SA, Scheuermann RH, Schober D, Smith B, Soldatova LN, Stoeckert CJ Jr, Taylor CF, Torniai C, Turner JA, Vita R, Whetzel PL, Zheng J. The Ontology for Biomedical Investigations. *PLoS One*. 2016 Apr 29;11(4):e0154556. doi: 10.1371/journal.pone.0154556. eCollection 2016. PubMed PMID: 27128319; PubMed Central PMCID: PMC4851331.
2. Bresciani A, Paul S, Schommer N, Dillon MB, Bancroft T, Greenbaum J, Sette A, Nielsen M, Peters B. T-cell recognition is shaped by epitope sequence conservation in the host proteome and microbiome. *Immunology*. 2016 May;148(1):34-9. doi: 10.1111/imm.12585. Epub 2016 Feb 8. PubMed PMID: 26789414; PubMed Central PMCID: PMC4819143.
3. Fleri W, Salimi N, Vita R, Peters B, Sette A, Immune Epitope Database and Analysis Resource, Encyclopedia of Immunobiology, Vol. 2, Elsevier Ltd., 20 May 2016, doi:10.1016/B978-0-12-374279-7.06004-5
4. McMurtrey C, Trolle T, Sansom T, Remesh SG, Kaever T, Bardet W, Jackson K, McLeod R, Sette A, Nielsen M, Zajonc DM, Blader IJ, Peters B, Hildebrand W. Toxoplasma gondii peptide ligands open the gate of the HLA class I binding groove. *Elife*. 2016 Jan 29;5. pii: e12556. doi: 10.7554/eLife.12556. PubMed PMID: 26824387; PubMed Central PMCID: PMC4775218.
5. Nielsen M, Andreatta M, NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length data sets. *Genome Medicine*, March 2016, Mar 30;8(1):33. doi: 10.1186/s13073-016-0288-x. PubMed PMID: 27029192; PubMed Central PMCID: PMC4812631
6. Paul S, Sidney J, Sette A, Peters B. TepiTool: A Pipeline for Computational Prediction of T Cell Epitope Candidates. *Curr Protoc Immunol*. 2016 Aug 1;114:18.19.1-18.19.24. doi: 10.1002/cpim.12. PubMed PMID: 27479659; PubMed Central PMCID: PMC4981331
7. Ponomarenko J, Vaughan K, Paul S, Peters B, Sette A, Haeussler M, Maurer-Stroh S, Ebola: an analysis of immunity at the molecular level, 2015 International Workshop on Artificial Immune

Systems (AIS), Taormina, 2015, pp. 1-8. doi: 10.1109/AISW.2015.7469230, published online at IEEE Xplore Digital Library, URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7469230&isnumber=7469223>

8. Trolle T, McMurtrey CP, Sidney J, Bardet W, Osborn SC, Kaeffer T, Sette A, Hildebrand WH, Nielsen M, Peters B. The Length Distribution of Class I-Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference. *J Immunol.* 2016 Feb 15;196(4):1480-7. doi: 10.4049/jimmunol.1501721. Epub 2016 Jan 18. PubMed PMID: 26783342; PubMed Central PMCID: PMC4744552.
9. Vita R, Overton JA, Seymour E, Sidney J, Kaufman J, Tallmadge RL, Ellis S, Hammond J, Butcher GW, Sette A, Peters B. An ontology for major histocompatibility restriction. *J Biomed Semantics.* 2016 Jan 11;7:1. doi:10.1186/s13326-016-0045-5. eCollection 2016. PubMed PMID: 26759709; PubMed Central PMCID: PMC4709943.
10. Xu X, Vaughan K, Weiskopf D, Grifoni A, Diamond MS, Sette A, Peters B. Identifying Candidate Targets of Immune Responses in Zika Virus Based on Homology to Epitopes in Other Flavivirus Species. *PLoS Curr.* 2016 Nov 15;8. pii: ecurrents.outbreaks.9aa2e1fb61b0f632f58a098773008c4b. doi: 10.1371/currents.outbreaks.9aa2e1fb61b0f632f58a098773008c4b. PubMed PMID: 28018746; PubMed Central PMCID: PMC5145810.

### **3.2 2015**

11. Trolle T, Metushi IG, Greenbaum JA, Kim Y, Sidney J, Lund O, Sette A, Peters B, Nielsen M. “Automated benchmarking of peptide-MHC class I binding predictions”. *Bioinformatics.* 2015 Feb 25. pii: btv123. [Epub ahead of print] PubMed PMID: 25717196
12. Vita, R. The Much Overlooked “Materials and Methods” (Editorial), *Immunome Research*, March 23, 2015 (<http://www.omicsonline.com/open-access/the-much-overlooked-materials-and-methods-1745-7580-1000e004.pdf>)
13. Sela-Culang, I., 1, Ofran, Y., and Peters, B. “Antibody specific epitope prediction - emergence of a new paradigm”, *Current Opinion in Virology*, Volume 11, April 2015, Pages 98–102, PMID: 25837466
14. Paul S, Lindestam Arlehamn CS, Scriba TJ, Dillon MB, Oseroff C, Hinz D, McKinney DM, Carrasco Pro S, Sidney J, Peters B, Sette A. Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *J Immunol Methods.* 2015 Jul;422:28-34. doi: 10.1016/j.jim.2015.03.022. Epub 2015 Apr 7. PubMed PMID: 25862607
15. Paul S, Dillon MB, Lindestam Arlehamn CS, Huang H, Davis MM, McKinney DM, Scriba TJ, Sidney J, Peters B, Sette A. A Population Response Analysis Approach To Assign Class II HLA-Epitope Restrictions. *J Immunol.* 2015 Jun 15;194(12):6164-76. doi: 10.4049/jimmunol.1403074. Epub 2015 May 6. PubMed PMID: 25948811; PubMed Central PMCID: PMC4458389
16. Vaughan K, Ponomarenko J, Peters B, Sette A. Analysis of Human RSV Immunity at the Molecular Level: Learning from the Past and Present. *PLoS One.* 2015 May 22;10(5):e0127108. doi: 10.1371/journal.pone.0127108. eCollection 2015. PubMed PMID: 26001197
17. Klausen MS, Anderson MV, Jespersen MC, Nielsen M, Marcatili P. LYRA, a webserver for lymphocyte receptor structural modeling. *Nucleic Acids Res.* 2015 Jul 1;43(W1):W349-W355. Epub 2015 May 24. PubMed PMID: 26007650

18. Andreatta M, Karosiene E, Rasmussen M, Stryhn A, Buus S, Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics*. 2015 Sep 29. [Epub ahead of print] PubMed PMID: 26416257
19. Carrasco Pro S, Sidney J, Paul S, Lindestam Arlehamn C, Weiskopf D, Peters B, Sette A. Automatic Generation of Validated Specific Epitope Sets. *J Immunol Res*. 2015;2015:763461. doi: 10.1155/2015/763461. Epub 2015 Oct 19. PubMed PMID: 26568965
20. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*. 2015 Oct 29. pii: btv639. [Epub ahead of print] PubMed PMID: 26515819
21. Sette A, Paul S, Vaughan K, Peters B. The Use of the Immune Epitope Database to Study Autoimmune Epitope Data Related to Alopecia Areata. *J Investig Dermatol Symp Proc*. 2015 Nov;17(2):36-41. doi: 10.1038/jidsymp.2015.39. Review. PubMed PMID: 26551944
22. Vita R, Vasilevsky N, Bandrowski A, Haendel M, Sette A, Peters B. Reproducibility and Conflicts in Immune Epitope Data. *Immunology*. 2015 Dec 17. doi: 10.1111/imm.12566. [Epub ahead of print] PubMed PMID: 26678806

### **3.3 2014**

23. Ponomarenko J, Vaughan K, Sette A, Maurer-Stroh S. Conservancy of mAb Epitopes in Ebolavirus Glycoproteins of Previous and 2014 Outbreaks. *PLOS Currents Outbreaks*. 2014 Nov 3. Edition 1.
24. Vaughan K, Peters B, Mallone R, von Herrath M, Roep BO, Sette A. Navigating diabetes-related immune epitope data: resources and tools provided by the Immune Epitope Database (IEDB). *Immunome Res*. 2013;9(1). doi: 10.4172/1745-7580.1000063. PubMed PMID: 25140192; PubMed Central PMCID: PMC4134942.
25. Vaughan K, Seymour E, Peters B, Sette A. Substantial gaps in knowledge of Bordetella pertussis antibody and T cell epitopes relevant for natural immunity and vaccine efficacy. *Hum Immunol*. 2014 May;75(5):440-451. doi: 10.1016/j.humimm.2014.02.013. Epub 2014 Feb 12. Review. PubMed PMID: 24530743.
26. Trolle T, Nielsen M. NetTepi: an integrated method for the prediction of T cell epitopes. *Immunogenetics*. 2014 May 27. [Epub ahead of print] PubMed PMID: 24863339
27. Kim Y, Sidney J, Buus S, Sette A, Nielsen M, Peters B. Dataset size and composition impact the reliability of performance benchmarks for peptide-MHC binding predictions. *BMC Bioinformatics*. 2014 Jul 14;15:241. doi: 10.1186/1471-2105-15-241. PubMed PMID: 25017736; PubMed Central PMCID: PMC4111843.
28. Shirai H, Prades C, Vita R, Marcatili P, Popovic B, Xu J, Overington JP, Hirayama K, Soga S, Tsunoyama K, Clark D, Lefranc MP, Ikeda K. Antibody informatics for drug discovery. *Biochim Biophys Acta*. 2014 Nov;1844(11):2002-2015. doi: 10.1016/j.bbapap.2014.07.006. Epub 2014 Aug 8. Review. PubMed PMID: 25110827.
29. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A, Peters B. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res*. 2014 Oct 9. pii: gku938. [Epub ahead of print] PubMed PMID: 25300482.

### **3.4 2013**

30. Kim Y, Yewdell JW, Sette A, Peters B. Positional bias of MHC class I restricted T-cell epitopes in viral antigens is likely due to a bias in conservation. *PLoS Comput Biol*. 2013

Jan;9(1):e1002884. doi: 10.1371/journal.pcbi.1002884. Epub 2013 Jan 24. PubMed PMID: 23357871; PubMed Central PMCID: PMC3554532.

31. Vita R, Overton JA, Greenbaum JA, Sette A; OBI consortium, Peters B. Query enhancement through the practical application of ontology: the IEDB and OBI. *J Biomed Semantics*. 2013 Apr 15;4 Suppl 1:S6. doi: 10.1186/2041-1480-4-S1-S6. Epub 2013 Apr 15. PubMed PMID: 23734660; PubMed Central PMCID: PMC3633001.
32. Jessen LE, Hoof I, Lund O, Nielsen M. SigniSite: Identification of residue-level genotype-phenotype correlations in protein multiple sequence alignments. *Nucleic Acids Res*. 2013 Jul 1;41(Web Server issue):W286-91. doi: 10.1093/nar/gkt497. Epub 2013 Jun 12. PMID: 23761454; PMCID: PMC3692133
33. Jørgensen KW, Rasmussen M, Buus S, Nielsen M. NetMHCstab - predicting stability of peptide:MHC-I complexes; impacts for CTL epitope discovery. *Immunology*. 2013 Aug 8. doi: 10.1111/imm.12160. PubMed PMID: 23927693; PubMed Central PMCID: PMC3893846
34. Schubert B, Lund O, Nielsen M. Evaluation of peptide selection approaches for epitope-based vaccine design. *Tissue Antigens*. 2013 Oct;82(4):243-51. doi: 10.1111/tan.12199. PubMed PMID: 24461003
35. Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics*. 2013 Oct;65(10):711-24. doi: 10.1007/s00251-013-0720-y. Epub 2013 Jul 31. PubMed PMID: 23900783.
36. Vaughan K, Peters B, O'Connor KC, Martin R, Sette A. A molecular view of multiple sclerosis and experimental autoimmune encephalitis: What can we learn from the epitope data? *J Neuroimmunol*. 2013 Dec 12. pii: S0165-5728(13)00340-8. doi: 10.1016/j.jneuroim.2013.12.009. PubMed PMID: 24365494.
37. Paul S, Kolla RV, Sidney J, Weiskopf D, Fleri W, Kim Y, Peters B, Sette A. Evaluating the immunogenicity of protein drugs by applying in vitro MHC binding data and the immune epitope database and analysis resource. *Clin Dev Immunol*. 2013;2013:467852. doi: 10.1155/2013/467852. Epub 2013 Oct 8. PubMed PMID: 24222776; PubMed Central PMCID: PMC3816028.
38. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Kesmir C, Peters B. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol*. 2013 Oct;9(10):e1003266. doi: 10.1371/journal.pcbi.1003266. Epub 2013 Oct 24. PubMed PMID: 24204222; PubMed Central PMCID: PMC3808449.
39. Paul S, Weiskopf D, Angelo MA, Sidney J, Peters B, Sette A. HLA Class I Alleles Are Associated with Peptide-Binding Repertoires of Different Size, Affinity, and Immunogenicity. *J Immunol*. 2013 Dec 15;191(12):5831-9. doi: 10.4049/jimmunol.1302101. Epub 2013 Nov 4. PubMed PMID: 24190657; PubMed Central PMCID: PMC3872965.

### **3.5 2012**

40. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, Lundegaard C, Sette A, Lund O, Bourne PE, Nielsen M, Peters B. Immune epitope database analysis resource. *Nucleic Acids Res*. 2012 Jul;40 (Web Server issue):W525-30. Epub 2012 May 18. PubMed PMID: 22610854.
41. Kim Y, Vaughan K, Greenbaum J, Peters B, Law M, Sette A. A Meta-Analysis of the Existing Knowledge of Immunoreactivity against Hepatitis C Virus (HCV). *PLoS One*. 2012;7(5):e38028. Epub 2012 May 31. PubMed PMID: 22675428; PubMed Central PMCID: PMC3364976.

42. Kringelum JV, Nielsen M, Padkjær SB, Lund O. Structural analysis of B-cell epitopes in antibody:protein complexes. *Mol Immunol.* 2013 Jan;53(1-2):24-34. doi: 10.1016/j.molimm.2012.06.001. Epub 2012 Jul 10. PubMed PMID: 22784991; PubMed Central PMCID: PMC3461403
43. Salimi N, Fleri W, Peters B, Sette A. The Immune Epitope Database: A Historical Retrospective of the First Decade. *Immunology.* 2012 Jun 8. doi:10.1111/j.1365-2567.2012.03611.x. [Epub ahead of print] PubMed PMID: 22681406.
44. Vaughan K, Kim Y, Sette A. A comparison of epitope repertoires associated with myasthenia gravis in humans and nonhuman hosts. *Autoimmune Dis.* 2012;2012:403915. doi: 10.1155/2012/403915. Epub 2012 Dec 2. PubMed PMID: 23243503; PubMed Central PMCID: PMC3518085.
45. Vaughan K, Peters B, Larche M, Pomes A, Broide D, Sette A. Strategies to Query and Display Allergy-Derived Epitope Data from the Immune Epitope Database. *Int Arch Allergy Immunol.* 2012 Nov 21;160(4):334-345. [Epub ahead of print] PubMed PMID: 23172234.

### **3.6 2011**

46. Seymour E., Damle R., Sette A., Peters B. (2011). "Cost sensitive hierarchical document classification to triage PubMed abstracts for manual curation." *BMC Bioinformatics* 12(1): 482. [Epub ahead of print]. PMID: 22182279.
47. Vaughan K, Greenbaum J, Kim Y, Vita R, Chung J, Peters B, Broide D, Goodman R, Grey H, Sette A. Towards defining molecular determinants recognized by adaptive immunity in allergic disease: an inventory of the available data. *J Allergy (Cairo).* 2010;2010:628026. Epub 2011 Feb 13. PubMed PMID: 21403821
48. Vita R, Peters B, Josephs Z, de Matos P, Ennis M, Turner S, Steinbeck C, Seymour E, Zarebski L, Sette A. (2011). "A Model for Collaborative Curation, The IEDB and ChEBI Curation of Non-peptidic Epitopes." *Immunome Res.* 2011 Apr 1;7(1):1-8. PubMed PMID: 21897450

### **3.7 2010**

49. Brinkman RR, Courtot M, Derom D, Fostel JM, He Y, Lord P, Malone J, Parkinson H, Peters B, Rocca-Serra P, Ruttenger A, Sansone SA, Soldatova LN, Stoeckert CJ Jr, Turner JA, Zheng J; OBI consortium. Modeling biomedical experimental processes with OBI. *J Biomed Semantics.* 2010 Jun 22;1 Suppl 1:S7. PubMed PMID: 20626927
50. Diehl AD, Augustine AD, Blake JA, Cowell LG, Gold ES, Gondré-Lewis TA, Masci AM, Meehan TF, Morel PA, Nijnik A, Peters B, Pulendran B, Scheuermann RH, Yao QA, Zand MS, Mungall CJ. "Hematopoietic cell types: Prototype for a revised cell ontology. *J Biomed Inform.* 2010 Feb 1. [Epub ahead of print] PMID: 20123131
51. Kim Y, Sette A, Peters B. Applications for T-cell epitope queries and tools in the Immune Epitope Database and Analysis Resource. *J Immunol Methods.* 2010 Oct 31. [Epub ahead of print] PubMed PMID: 21047510.
52. Nielsen M, Lund O, Buus S, Lundegaard C. MHC class II epitope predictive algorithms. *Immunology.* 2010 Jul;130(3):319-28. Epub 2010 Apr 12. Review. PubMed PMID: 20408898;
53. Ponomarenko J, Papangelopoulos N, Zajonc DM, Peters B, Sette A, Bourne PE. IEDB-3D: structural data within the immune epitope database. *Nucleic Acids Res.* 2010 Oct 28. [Epub ahead of print] PubMed PMID: 21030437.

54. Salimi N, Fleri W, Peters B, Sette A. Design and utilization of epitope-based databases and predictive tools. *Immunogenetics*. 2010 Apr;62(4):185-96. Epub 2010 Mar 6. Review. PubMed PMID: 20213141; PubMed Central PMCID: PMC2843836.
55. Vaughan K, Greenbaum J, Blythe M, Peters B, Sette A. Meta-analysis of all immune epitope data in the Flavivirus genus: inventory of current immune epitope data status in the context of virus immunity and immunopathology. *Viral Immunol*. 2010 Jun;23(3):259-84. PubMed PMID: 20565291; PubMed Central PMCID: PMC2942863.
56. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*. 2010 Nov 22;11:568. PubMed PMID: 21092157; PubMed Central PMCID: PMC2998531.
57. Zhang H, Wang P, Papangelopoulos N, Xu Y, Sette A, Bourne PE, Lund O, Ponomarenko J, Nielsen M, Peters B. Limitations of Ab initio predictions of peptide binding to MHC class II molecules. *PLoS One*. 2010 Feb 17;5(2):e9272. PubMed PMID: 20174654; PubMed Central PMCID: PMC2822856.

### **3.8 2009**

58. Davies V, Vaughan K, Damle R, Peters B, Sette A. Classification of the universe of immune epitope literature: representation and knowledge gaps. *PLoS One*. 2009 Sep 14;4(9):e6948. PubMed PMID: 19774228;
59. Greenbaum JA, Kotturi MF, Kim Y, Oseroff C, Vaughan K, Salimi N, Vita R, Ponomarenko J, Scheuermann RH, Sette A, Peters B. Pre-existing immunity against swine-origin H1N1 influenza viruses in the general human population. *Proc Natl Acad Sci U S A*. 2009 Dec 1;106(48):20365-70. Epub 2009 Nov 16. PubMed PMID: 19918065; PubMed Central PMCID: PMC2777968.
60. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M, NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics*. 2009 Jan;61(1):1-13. PMID: 19002680
61. Kim Y, Sidney J, Pinilla C, Sette A, Peters B. Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. *BMC Bioinformatics*. 2009 Nov 30;10:394. PubMed PMID: 19948066.
62. Vaughan K, Blythe M, Greenbaum J, Zhang Q, Peters B, Doolan DL, Sette A, Meta-analysis of immune epitope data for all Plasmodia: overview and applications for malarial immunobiology and vaccine-related issues. *Parasite Immunol*. 2009 Feb;31(2):78-97. [PMID: 19149776]
63. Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B. The Immune Epitope Database 2.0. *Nucleic Acids Res*. 2009 Nov 11. [Epub ahead of print] PubMed PMID: 19906713.
64. Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: Application to MHC-peptide binding. *Bioinformatics*. 2009 Mar 17. PMID: 19297351

### **3.9 2008**

65. Lundegaard C., K. Lamberth, M. Harndahl, S. Buus, O. Lund, M. Nielsen. (2008). "NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11." *Nucleic Acids Res*. 2008 36:W509-12. PMID: 18996943

66. Nielsen, M., C. Lundegaard, T. Blicher, B. Peters, A. Sette, S. Justesen, S. Buus, O. Lund. (2008). "Quantitative Predictions of Peptide Binding to Any HLA-DR Molecule of Known Sequence: NetMHCIIpan." *PLoS Computational Biology* **4**(7). PMID: 18604266
67. Ponomarenko J, H. H. Bui, W. Li, N. Fusseder, P. E. Bourne, A. Sette, B. Peters. (2008). "ElliPro: a new structure-based tool for the prediction of antibody epitopes." *BMC Bioinformatics*, **9**(514). PMID: 19055730
68. Ponomarenko J. V., M. H. C. van Regenmortel (2008) "B-cell epitope prediction." *Structural Bioinformatics*, Ed: Bourne PE, Gu J. Wiley-Liss; 2 edition.
69. Rapin N., I. Hoof, O. Lund, M. Nielsen. (2008). "MHC motif viewer." *Immunogenetics*. **60**(12): 759-65. PMID: 18766337
70. Vita R., B. Peters, A. Sette. (2008). "The Curation Guidelines of the Immune Epitope Database and Analysis Resource." *Cytometry A* **73**(11): 1066-1070. PMID: 18688821
71. Wang, P., J. Sidney, C. Dow, B. Mothe, A. Sette, B. Peters. (2008). "A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach." *PLoS Computational Biology* **4**(4). PMID: 18389056
72. Zarebski L. M., K. Vaughan, J. Sidney, B. Peters, H. Grey, K. D. Janda, A. Casadevall, A. Sette. (2008). "Analysis of epitope information related to *Bacillus anthracis* and *Clostridium botulinum*." *Expert Rev Vaccines* **7**(1): 55-74. PMID: 18251694
73. Zhang, Q., P. Wang, et al. (2008). "Immune epitope database analysis resource (IEDB-AR)." *Nucleic Acids Research* **36**: W513-W518. PMID: 18515843

### **3.102007**

74. Beaver JE, Bourne PE, Ponomarenko JV. EpitopeViewer: a Java application for the visualization and analysis of immune epitopes in the Immune Epitope Database and Analysis Resource (IEDB). *Immunome Res.* 2007 Feb 21;3:3. PMID: 17313688
75. Blythe MJ, Zhang Q, Vaughan K, de Castro R Jr, Salimi N, Bui HH, Lewinsohn DM, Ernst JD, Peters B, Sette A. An analysis of the epitope knowledge related to Mycobacteria. *Immunome Res.* 2007 Dec 14;3(1):10. PMID: 18081934
76. Bui HH, Peters B, Assarsson E, Mbawuikie I, Sette A. Ab and T cell epitopes of influenza A virus, knowledge and opportunities. *Proc Natl Acad Sci U S A.* 2007 Jan 2;104(1):246-51. PMID: 17200302
77. Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics.* 2007 Sep 26;8(1):361. PMID: 17897458
78. Ernst JD, Lewinsohn DM, Behar S, Blythe M, Schlesinger LS, Kornfeld H, Sette A. Meeting Report: NIH Workshop on the Tuberculosis Immune Epitope Database. *Tuberculosis (Edinb).* 2007 Dec 6. PMID: 18068490
79. Greenbaum JA, Andersen PH, Blythe M, Bui HH, Cachau RE, Crowe J, Davies M, Kolaskar AS, Lund O, Morrison S, Mumey B, Ofra Y, Pellequer JL, Pinilla C, Ponomarenko JV, Raghava GP, van Regenmortel MH, Roggen EL, Sette A, Schlessinger A, Sollner J, Zand M, Peters B. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J Mol Recognit.* 2007 Mar-Apr;20(2):75-82. PMID: 17205610
80. Lundegaard C, Lund O, Kesmir C, Brunak S, Nielsen M. Modeling the adaptive immune system: predictions and simulations. *Bioinformatics.* 2007 Dec 15;23(24):3265-75. PMID: 18045832

81. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, Røder G, Peters B, Sette A, Lund O, Buus S. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE*. 2007 Aug 29;2(8):e796. PMID: 17726526
82. Peters B, Sette A. Integrating epitope data into the emerging web of biomedical knowledge resources. *Nat Rev Immunol*. 2007 Jun;7(6):485-90. PMID: 17479127
83. Ponomarenko JV, Bourne PE., Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *BMC Struct Biol*. 2007 Oct 2;7(1):64. PMID: 17910770
84. Sette A, Peters B., Immune epitope mapping in the post-genomic era: lessons for vaccine development. *Curr Opin Immunol*. 2007 Feb;19(1):106-10. PMID: 17113275
85. Wang P, Morgan AA, Zhang Q, Sette A, Peters B. Automating document classification for the Immune Epitope Database. *BMC Bioinformatics*. 2007 Jul 26;8:269. PMID: 17655769

### **3.112006**

86. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2006 Mar 17;7(1):153. PMID: 16545123
87. Peters B, Bui HH, Frankild S, Nielson M, Lundegaard C, Kostem E, Basch D, Lamberth K, Harndahl M, Fleri W, Wilson SS, Sidney J, Lund O, Buus S, Sette A., A community resource benchmarking predictions of peptide binding to MHC-I molecules, *PLoS Comput Biol*. 2006 Jun 9;2(6):e65. Epub 2006 Jun 9. PMID: 16789818
88. Salimi N, R. Vita. (2006). "The biocurator: connecting and enhancing scientific data". *PLoS Comput Biol* 2(10) :e125. PMID: 17069454
89. Sette, A., H. H. Bui, S. Buus, W. Fleri, R. Kubo, O. Lund, D. Nemazee, J. V. Ponomarenko, M. Sathiamurthy, S. Stewart, S. Way, S. S. Wilson, B. Peters. (2006). "The Immune Epitope Database and Analysis Resource." *Pattern Recognition in Bioinformatics: International Workshop, Hong Kong, China, August 20, 2006 : Proceedings*, Springer, Jagath C. Rajapakse, Limsoon Wong, Raj Acharya (Eds). ISBN: 3540374469 9783540374466
90. Vita R, K. Vaughan, L. Zarebski, N. Salimi, W. Fleri, H. Grey, M. Sathiamurthy, J. Mokili, H. H. Bui, P. E. Bourne, J. V. Ponomarenko, R. de Castro Jr, R. K. Chan, J. Sidney, S. S. Wilson, S. Stewart, S. Way, B. Peters, A. Sette. (2006) "Curation of complex, context-dependent immunological data." *BMC Bioinformatics*. 12(7): 341. PMID: 16836764

### **3.122005**

91. Bui HH, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton KA, Mothé BR, Chisari FV, Watkins DI, Sette A. Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics*. 2005 Jun;57(5):304-14. Epub 2005 May 3. PMID: 15868141
92. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics*. 2005 May 31;6:132. PMID: 15927070
93. Peters, B., J. Sidney, P. Bourne, H. H. Bui, S. Buus, G. Doh, W. Fleri, M. Kronenberg, R. Kubo, O. Lund, D. Nemazee, J. V. Ponomarenko, M. Sathiamurthy, S. Schoenberger, S. Stewart, P. Surko, S. Way, S. Wilson, A. Sette. (2005). "The design and implementation of the immune epitope database and analysis resource." *Immunogenetics* 57(5): 326-336. PMID: 15895191

94. Peters, B., J. Sidney, P. Bourne, H. H. Bui, S. Buus, G. Doh, W. Fleri, M. Kronenberg, R. Kubo, O. Lund, D. Nemazee, J. V. Ponomarenko, M. Sathiamurthy, S. Schoenberger, S. Stewart, P. Surko, S. Way, S. Wilson, A. Sette. (2005). "The Immune Epitope Database and Analysis Resource: From Vision to Blueprint." *PLoS Biology* 3(3). PMID: 15760272
95. Sathiamurthy, M., B. Peters, H. H. Bui, J. Sidney, J. Mokili, S. S. Wilson, W. Fleri, D. McGuinness, P. Bourne, A. Sette. (2005). "An ontology for immune epitopes: application to the design of a broad scope database of immune reactivities." *Immunome Res* 1(2): 1745-7580. PMID: 16305755
96. Sette, A., W. Fleri, B. Peters, M. Sathiamurthy, H. H. Bui, S. Wilson. (2005). "A Roadmap for the Immunomics of Category A–C Pathogens." *Immunity* 22(2): 155-161. PMID: 15773067

### **3.132004**

97. Sette A. (2004). "The immune epitope database and analysis resource: from vision to blueprint." *Genome Inform Ser Workshop Genome Inform* 15(2):299. PMID: 16312048

## 4 References

- Andreatta M, Karosiene E, Rasmussen M, Stryhn A, Buus S, Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics*. 2015 Nov;67(11-12):641-50. doi: 10.1007/s00251-015-0873-y. Epub 2015 Sep 29. PubMed PMID: 26416257; PubMed Central PMCID: PMC4637192.
- Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*. 2016 Feb 15;32(4):511-7. doi: 10.1093/bioinformatics/btv639. Epub 2015 Oct 29. PubMed PMID: 26515819.
- Bui HH, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton KA, Mothe BR, Chisari FV, Watkins DI, Sette A. Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics*. 2005 Jun;57(5):304-14. Epub 2005 May 3.
- Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2006 Mar 17;7(1):153. PMID: 16545123
- Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2007 Sep 26;8(1):361. PMID: 17897458
- Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Keşmir C, Peters B. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol*. 2013 Oct;9(10):e1003266. doi:10.1371/journal.pcbi.1003266. Epub 2013 Oct 24. PubMed PMID: 24204222; PubMed Central PMCID: PMC3808449.
- Chou P, Fasman G. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol*. 1978;45-148. PMID: 364941
- Emini E, Hughes J, Perlow D, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus specific synthetic peptide. *J Virol*. 1985;55:836-839. PMID: 2991600
- Giguère S, Drouin A, Lacoste A, Marchand M, Corbeil J, Laviolette F. MHC-NP: predicting peptides naturally processed by the MHC. *J Immunol Methods*. 2013 Dec 31;400-401:30-6. doi: 10.1016/j.jim.2013.10.003. Epub 2013 Oct 18. PubMed PMID: 24144535.
- González-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MH, da Silva AL, Teles e Silva AL, Ghataoraya GS, Alfirevic A, Jones AR, Middleton D. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res*. 2015 Jan;43(Database issue):D784-8. doi: 10.1093/nar/gku1166. Epub 2014 Nov 20. PubMed PMID: 25414323; PubMed Central PMCID: PMC4383964
- Haste Andersen P, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci*. 2006 Nov;15(11):2558-67. Epub 2006 Sep 25. PMID: 17001032
- Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M, NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics*. 2009 Jan;61(1):1-13. PMID: 19002680
- Karosiene E, Lundegaard C, Lund O and Nielsen M. 2012. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* **64(3)**:177-186. [PMID: 22009319](https://pubmed.ncbi.nlm.nih.gov/22009319/)

- Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. *Naturwissenschaften*. 1985. 72:212-213.
- Keşmir C, Nussbaum AK, Schild H, Detours V, Brunak S. Prediction of proteasome cleavage motifs by neural networks. *Protein Eng*. 2002 Apr;15(4):287-96. PMID: 11983929
- Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, Lundegaard C, Sette A, Lund O, Bourne PE, Nielsen M, Peters B. Immune epitope database analysis resource. *Nucleic Acids Res*. 2012 Jul;40(Web Server issue):W525-30. doi: 10.1093/nar/gks438. Epub 2012 May 18. PubMed PMID: 22610854; PubMed Central PMCID: PMC3394288
- Kim Y, Sidney J, Buus S, Sette A, Nielsen M, Peters B. Dataset size and composition impact the reliability of performance benchmarks for peptide-MHC binding predictions. *BMC Bioinformatics*. 2014 Jul 14;15:241. doi: 10.1186/1471-2105-15-241. PubMed PMID: 25017736; PubMed Central PMCID: PMC4111843
- Kim Y, Sidney J, Pinilla C, Sette A, Peters B. Derivation of an amino acid similarity matrix for peptide-MHC binding and its application as a Bayesian prior. *BMC Bioinformatics*. 2009 Nov 30;10:394. PMID: 19948066
- Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*. 1990 Dec 10;276(1-2):172-4. PMID: 1702393
- Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol*. 2012;8(12):e1002829. doi: 10.1371/journal.pcbi.1002829. Epub 2012 Dec 27. PubMed PMID: 23300419; PubMed Central PMCID: PMC3531324.
- Kringelum JV, Nielsen M, Padkjær SB, Lund O. Structural analysis of B-cell epitopes in antibody:protein complexes. *Mol Immunol*. 2013 Jan;53(1-2):24-34. doi: 10.1016/j.molimm.2012.06.001. Epub 2012 Jul 10. PubMed PMID: 22784991; PubMed Central PMCID: PMC3461403
- Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res*. 2006 Apr 24;2:2. PMID: 16635264
- Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, Nielsen M. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol*. 2005 Aug;35(8):2295-303. PMID: 15997466
- Liang S, Zheng D, Standley DM, Yao B, Zacharias M, Zhang C. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics*. 2010 Jul 16;11:381. doi: 10.1186/1471-2105-11-381. PubMed PMID: 20637083; PubMed Central PMCID: PMC2910724.
- Marcatili P, Rosi A, Tramontano A., 2008. PIGS: automatic prediction of antibody structures. *Bioinformatics* 24(17):1953-4. PMID: 18641403
- Moutaftsi M, Peters B, Pasquetto V, Tschärke DC, Sidney J, Bui HH, Grey H, Sette A. A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat Biotechnol*. 2006 Jul;24(7):817-9. PMID: 16767078
- Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med*. 2016 Mar 30;8(1):33. doi: 10.1186/s13073-016-0288-x. PubMed PMID: 27029192; PubMed Central PMCID: PMC4812631.

- Nielsen M, Lundegaard C, Blicher T, Peters B, Sette A, Justesen S, Buus S, and Lund O. 2008. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput Biol.* 4(7):e1000107. PMID: 18604266
- Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics.* 2007 Jul 4;8:238. PMID: 17608956
- Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, Brunak S, Lund O., Reliable prediction of T-cell epitopes using neural networks with novel sequence representations, *Protein Sci.* 2003 May;12(5):1007-17
- 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;25:5425–5432. PMID: 2430611
- Paul S, Dillon MB, Lindestam Arlehamn CS, Huang H, Davis MM, McKinney DM, Scriba TJ, Sidney J, Peters B, Sette A. A population response analysis approach to assign class II HLA-epitope restrictions. *J Immunol.* 2015 Jun 15;194(12):6164-76. doi: 10.4049/jimmunol.1403074. Epub 2015 May 6. PubMed PMID: 25948811; PubMed Central PMCID: PMC4458389.
- Paul S, Sidney J, Sette A, Peters B. TepiTool: A Pipeline for Computational Prediction of T Cell Epitope Candidates. *Curr Protoc Immunol.* 2016 Aug 1;114:18.19.1-18.19.24. doi: 10.1002/cpim.12. PubMed PMID: 27479659; PubMed Central PMCID: PMC4981331.
- Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, Fleri W, Kronenberg M, Kubo R, Lund O, Nemazee D, Ponomarenko JV, Sathiamurthy M, Schoenberger S, Stewart S, Surko P, Way S, Wilson S, Sette A. The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol.* 2005 Mar;3(3):e91. PMID: 15760272.
- Peters B, Sette A., "Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method." *BMC Bioinformatics* 2005 May 31;6(1):132
- Peters B, Bulik S, Tampe R, Van Endert PM, Holzhtutter HG. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J Immunol.* 2003 Aug 15;171(4):1741-9.
- Ponomarenko JV, Bourne PE., Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *BMC Struct Biol.* 2007 Oct 2;7(1):64. PMID: 17910770
- Ponomarenko J, H. H. Bui, W. Li, N. Fusseder, P. E. Bourne, A. Sette, B. Peters. (2008). "ElliPro: a new structure-based tool for the prediction of antibody epitopes." *BMC Bioinformatics*, 9(514). PMID: 19055730
- Rasmussen M, Fenoy E, Harndahl M, Kristensen AB, Nielsen IK, Nielsen M, Buus S. Pan-Specific Prediction of Peptide-MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity. *J Immunol.* 2016 Aug 15;197(4):1517-24. doi: 10.4049/jimmunol.1600582. Epub 2016 Jul 8. PubMed PMID: 27402703; PubMed Central PMCID: PMC4976001.
- Rubinstein ND, Mayrose I, Martz E, Pupko T. Epitopia: a web-server for predicting B-cell epitopes. *BMC Bioinformatics.* 2009 Sep 14;10:287. doi: 10.1186/1471-2105-10-287. PubMed PMID: 19751513; PubMed Central PMCID: PMC2751785.
- Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, Peters B. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Res.* 2008 Jan 25;4:2. PMID: 18221540

- Stranzl T, Larsen MV, Lundegaard C, Nielsen M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics*. 2010 Jun;62(6):357-68. Epub 2010 Apr 9. PubMed PMID: 20379710; PubMed Central PMCID: PMC2875469
- Sturniolo T, Bono E, Ding J, Radrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol*. 1999 Jun;17(6):555-61. PMID: 10385319
- Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel PM, Rammensee HG, Schild H, Holzhutter HG. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci*. 2005 May;62(9):1025-37.
- Trolle T, Metushi IG, Greenbaum JA, Kim Y, Sidney J, Lund O, Sette A, Peters B, Nielsen M. Automated benchmarking of peptide-MHC class I binding predictions. *Bioinformatics*. 2015 Jul 1;31(13):2174-81. doi: 10.1093/bioinformatics/btv123. Epub 2015 Feb 25. PubMed PMID: 25717196; PubMed Central PMCID: PMC4481849.
- Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A, Peters B. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res*. 2014 Oct 9. pii: gku938. [Epub ahead of print] PubMed PMID: [25300482](#).
- Wang, P., J. Sidney, C. Dow, B. Mothe, A. Sette, B. Peters. (2008). "A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach." *PLoS Computational Biology* 4(4). PMID: 18389056
- Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B. 2010. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*. 11:568
- Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: Application to MHC-peptide binding. *Bioinformatics*. 2009 Mar 17. PMID: 19297351