Contract No. HHSN266200400006C

Immune Epitope Database and Analysis Program

2005 Annual Compendium

La Jolla Institute for Allergy and Immunology
3030 Bunker Hill Street, Suite 326
San Diego, CA 92109

March 2006

Table of Contents

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. This program has an emphasis on epitopes from the NIAID Category A, B, and C Priority Pathogens, such as anthrax, smallpox, West Nile virus, influenza and SARS. The IEDB contains detailed information from humans, non-human primates, rodents, and any species for which detailed immune response information is available. The program includes the development and maintenance of analytical and predictive tools and methods.

This first Annual Compendium of the Immune Epitope Database and Analysis Resource consists of three sections. The first section lists and describes the various features of the IEDB website implemented by the end of 2005. The second section contains a list of the antibody and T cell epitope information in the database as of 31 December 2005. Although the IEDB collects data from three primary sources – published literature, patents, and direct data submissions from researcher – only literature data were curated in 2005. The third section lists the scientific publications in 2005 for which the IEDB played a contributory role. This list is limited to articles published by IEDB contractor team members given that it was not publicly available in 2005.

1 Website Features

The IEDB website functionality can be divided into four categories – Query, Resources, Account Information, and Support, 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. These features can also be classified by those available to any user and those available only to registered users. Any user can become a registered user in a simple process described later. The features are listed in Table 1.1

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, patent, or submission) containing information about one or more epitopes and associated binding or response information. The same epitope can exist in multiple references. Therefore, the results of a query can contain the same epitope numerous times if the epitope is contained in multiple references.

Anonymous User Features	Registered User Features
• Tour the IEDB • Submit feedback • View online help • View home page • Perform an Advanced Query • Perform a Simple Query • Perform a Search • Download Advanced Query results • Download Simple Query or Search results • Utilize analysis tools • Utilize finders when performing queries • View forum discussion topics • Search forum discussion topics • Download export files • Download documents • View external links	All features for anonymous users PLUS: • Submit a help request to the help desk • Save advanced queries so they can be used at a later time • Submit a discussion topic to a forum • Reply to a previously submitted discussion topic in a forum

Table 1.1 IEDB website features available to anonymous and registered users

1.1 Query

1.1.1 Query Overview

All users will be able to perform the following three types of queries: Search, Simple Query, and Advanced Query. When performing any type of query, users will be able to retrieve information from literature, patents, or submissions. Submissions contain relevant unpublished and unpatented epitope data.

All three types of queries are used to retrieve results with specified criteria. Criteria entered for any type of query are case insensitive. For example, "ABC", "abc", and "Abc" are all the same. A Search only searches selected fields in the IEDB. The Search feature uses one search field which is displayed in the upper right-hand corner of most pages of the application and only takes keywords, certain identifiers, or an epitope linear sequence as input. If the entered criteria are contained in one of the searchable fields of a record, the record will be included in the results.

The Simple and Advanced Queries make use of several different finders. These finders are used to select data item values among the IEDB-available controlled vocabulary. For example, the Species Finder lets the user search for and select species and strains from the NCBI taxonomy. Finders are available to facilitate the selection of allele, assay, disease, source, and species.

A Simple Query is designed for very common queries and has a limited number of search fields. This is the only search where users can specify epitope criteria, reference criteria, and criteria for more than one of the assay categories (MHC Binding, MHC Ligand Elution, T-Cell Response, and B-Cell Response) at once. The Species Finder and Allele Finder can be utilized when performing a Simple Query.

The Advanced Query is based on the Query By Example approach. Query By Example is an alias for Advanced Query so the terms are used interchangeably. The Advanced Query utilizes hundreds of fields and is the most complex type of search. With this type of search, epitope criteria, reference criteria, and criteria for at most one assay category can be entered. The Species Finder, Allele Finder, Assay Finder, and Disease Finder can be utilized when performing an Advanced Query.

Results for the Advanced Query are displayed differently from the results for the Search and Simple Query. The Advanced Query returns all assays and associated epitopes with the given search criteria as long as the entered information is contained in one reference (article, patent, or submission). Each reference contains information about one or more epitopes and associated binding or response information. The same epitope can exist in multiple references. Therefore, the results of an Advanced Query can contain the same epitope and associated assay information numerous times if the epitope is contained in multiple references. The results of the Search and Simple Query additionally include the distinct epitopes matching the search criteria. A distinct epitope is an epitope with given characteristics regardless of the reference. If the same epitope is described in two different references, and a user defines criteria that matches the epitope, only one distinct epitope will be displayed in the results.

1.1.2 Search and Simple Query – Results Summary Page

The Result Summary page displays the results of a Search or Simple Query. The number of references, records, distinct epitopes, distinct structures, and assays (displayed as links) that met the specified search criteria are listed.

The number of **references** is the total number of references containing the matching records. Each **record** contains data about one epitope (or structure) in one reference. A protein/DNA sequence/carbohydrate etc. that may or may not induce an immune response or MHC binding is considered a structure. A structure is an epitope if it produces a positive qualitative measurement for at least one of its associated assays. A structure is not an epitope if it produces a negative qualitative measurement for each of its associated assays. All epitopes are structures, but not all structures are epitopes. Multiple epitopes/structures can be described in a single reference. If the same epitope is described in two references, and a user defines criteria that match the epitope, two records would be returned which are identical except for the reference information.

A **distinct epitope** is an epitope with given characteristics regardless of the reference. If the same epitope is described in two different references, and a user defines criteria that matches the epitope, only one distinct epitope will be displayed in the results. The number of distinct epitopes will always be less than or equal to the number of distinct structures.

A **distinct structure** is a structure with given characteristics regardless of the reference. If the same structure is described in two different references, and a user defines criteria that matches the structure, only one distinct structure will be displayed in the results. The number of records will always be greater than or equal to the number of distinct structures.

The distinct epitopes and distinct structures are categorized according to chemical type.

3

Assays are divided by assay category. Assays are associated with records. Therefore, if the same epitope and associated assay are described in two references, and a user enters search criteria for the epitope, two identical assays (with the exception of the reference information) will be listed.

Users can access an Assay List page in numerous ways after performing a Simple Query or a Search by clicking on certain links. Results on an Assay List page can be downloaded in Excel format by clicking on the Excel link at the bottom of the screen. All assays will be downloaded regardless of the page shown or how many assays are displayed per page. Additionally, the results can be analyzed.

1.1.3 Advanced Query

The Advanced Search has numerous additional features that are not available with the Search and Simple Query. Most of these are available to any anonymous user, but several are only available to registered users. Further information on these features is provided in the following subsections.

1.1.3.1 View Query Criteria

The criteria of a query will include the entered search criteria for the query. All users will have the ability to view the criteria of an Advanced Query after the query has been performed or after search criteria has been input into the Perform an Advanced Query form. Additionally, registered users will be able to view the criteria of their saved advanced queries.

1.1.3.2 Save Queries

Registered users can save a query after they have performed an Advanced Query. Users cannot save a query after performing a Search or a Simple Query. A saved query will contain all of the entered filter criteria for the query and the format selected to report the results. When users save a query, they are required to give it a search title so it can be found in the future. Users might want to save a query so they can perform the query again at a later date when additional epitopes from other references have been added to the database. Once a query has been saved, users can load a saved query and then perform the query, or delete a saved query. Saved queries will expire after a determined period of time.

1.1.3.3 View Advanced Query Results

All users can view the results of an Advanced Query. To view the results of an Advanced Query, users must first perform the query for which they wish to view the results. The View Results page is displayed after an Advanced Query is performed which includes a Results Table. The Results Table will include epitope records with a limited amount of information and contains the following five columns: Links, Reference, Structure, Source, and Assay. One row will be displayed for each assay that meets the specified search criteria. Aggregate values for each assay will populate each column of the Results Table, except for the Links column, when the Summary format is selected. Under the Links column, each row in the table will have a Details link and possibly an Epitope Viewer link. Users can click on a Details link to view more detailed result information for the selected row. When a user clicks on a Details link, all populated reference fields, all populated epitope fields, and all populated binding or response

fields for the corresponding assay will be displayed. The Epitope Viewer is an application for three dimensional viewing of receptor-antigen interactions and is available for records with interacting residue data.

Users can format the Results Table to include columns other than the default columns. Users have the option to select from several predetermined formats using the change format function, or to create a custom format. By default, the Summary format is used. Users also have the option to alter the number of rows displayed per page in the Results Table.

1.1.3.4 Download Advanced Query Results

All users can download the results of an Advanced Query in Excel format. To download the results of a query, users must first perform the query for which they wish to download the results. Whatever columns are displayed in the Results Table when the user selects the Excel link will be the columns that are downloaded.

1.1.3.5 Accessing the EpitopeViewer

All users are able to utilize the EpitopeViewer after an Advanced Query has been performed. The EpitopeViewer is an application for three dimensional viewing of receptor-antigen interactions that can be accessed from the View Results screen. The EpitopeViewer can be used with all assays that have receptor-antigen interaction data available.

1.1.4 Finders Overview

Several finders (Allele, Assay, Disease, Source, and Species) are available to help facilitate selections. The allele, assay, and source finders utilize tables and the disease and species finders utilize hierarchical trees. All finders can be utilized when performing an Advanced Query. The allele, source, and species finders can also be used when performing a Simple Query. Multiple selections can be made when utilizing finders during a query. Records will be returned that contain one of the entered values, or at least one of the entered values in the case of the MHC Types Present field.

The disease and species finders use wildcard characters by default on both ends of entered search criteria when the Search names button is selected. An exact match is found when criteria are entered and the Search ID's button is selected. The allele finder uses wildcard characters by default on both ends of criteria entered into the allele field. The source finder uses wildcard characters by default on both ends of criteria entered into the name or strain fields. All finder search fields can search on common strings such as 'the', 'and', 'a', 'or', 'but', etc.

1.1.4.1 Allele Finder

The allele finder is used to facilitate the selection of an allele. The allele finder displays alleles in a table. Searching capabilities is provided so users can quickly select an allele. The allele finder can be used with the following fields: MHC Allele and MHC Types Present.

Anyone can use the allele finder when performing an Advanced Query or a Simple Query. Only the alleles referenced in the IEDB will be displayed in the allele finder if accessed from the query application. When performing an Advanced Query, multiple alleles can be added to any MHC Allele field and/or any MHC Types Present field as long as the MHC Allele field and the

MHC Types Present field exist on one tab on the query form. If multiple alleles are added to an MHC Allele field, they will be semicolon-separated on the parent query form. The query results will include records that have one of the entered MHC Alleles for the corresponding section. If multiple alleles are added to an MHC Types Present field, they will be semicolon-separated on the parent query form. The query results will include records that have one or more of the entered MHC Types Present for the corresponding section. When performing a Simple Query, multiple alleles can be added to the MHC Restriction field.

1.1.4.2 Assay Finder

The assay finder is used to facilitate the selection of an assay to input into an Assay Type field. The assay finder will display assay types in a table. Searching capabilities will be provided so users can quickly select an assay type.

Anyone can use the assay finder when performing an Advanced Query. When performing an Advanced Query, multiple values can be added to the Assay Type field which will be semicolon-separated on the parent query form. The query results will include records that have one of the entered assay types for the corresponding section.

The application will determine which assay category (MHC Binding, T Cell Response, B Cell Response, or MHC Ligand Elution) was being utilized when the assay finder was accessed, and all the assays from that single category will be displayed in the assay finder.

1.1.4.3 Disease Finder

The disease finder is used to facilitate the selection of a disease state and input the selection into a Disease State field. The disease finder includes all diseases from The International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10). The disease finder displays diseases and ICD-10 codes in a hierarchical tree. The first level of the tree displays similar groups of diseases, and each additional level of the tree further breaks down the groups of diseases. Variations of each disease are not included. Searching capabilities are provided so users can quickly select a disease state.

Anyone can use the disease finder when performing an Advanced Query. When performing an Advanced Query, multiple values can be added to a Disease State field which will be semicolon-separated on the parent query form. The query results will include records that have one of the entered disease states for the corresponding section. If a user selects a general category of a disease, such as Anthrax, results will be returned with the disease state 'Anthrax [A22]' as well as the sub-categories of that disease state, such as Cutaneous anthrax [A22.0], Pulmonary anthrax [A22.1], Gastrointestinal anthrax [A22.2], Anthrax septicemia [A22.7], Other forms of anthrax [A22.8], and Anthrax unspecified [A22.9]. If a user selects a specific disease state, such as Pulmonary anthrax [A22.1], only the results with the disease state 'Pulmonary anthrax [A22.1]' will be returned.

1.1.4.4 Source Finder

The source finder is used to facilitate the selection of an epitope, antigen, or immunogen source from any of the following databases: GenBank, Swiss-Prot, or IEDB. A Source Finder button is displayed next to every Source Name field. The source finder displays sources in a table.

Searching capabilities are provided so users can quickly select a source. When a source is selected during querying, only the source name is populated on the query form.

Anyone can use the source finder when performing an Advanced Query or Simple Query. When querying, multiple values can be added using the source finder which will be semicolon-separated on the parent query form. The query results include records that have one of the entered source names for the corresponding section.

1.1.4.5 Species Finder

The species finder is used to facilitate the selection of a species or virus from the NCBI Taxonomy Database. The species finder displays species in a hierarchical tree and a list of common species. The first level of the tree contains all species, and additional levels break species down into like categories until a specific species is displayed. Searching capabilities are provided so users can quickly select a species. The breadth-first searching algorithm is used. The tree is searched first by level and then by the order in the tree within each level. Matches include items in the tree as well as synonyms. Synonyms are listed in the box that is displayed when you move the cursor over a species name. Not all entries have synonyms. An example of a virus that does have synonyms is the Influenza A Virus.

Anyone can use the species finder when performing an Advanced Query or a Simple Query. Only the species referenced in the IEDB will be contained in the species finder hierarchical tree if accessed from the query application. When performing an Advanced Query, a species finder button is displayed next to every Source Species or Species field. When performing a Simple Query, a species finder button is displayed next to the Host/Immunized Species field and the Epitope Source Species field. Multiple values can be added to these fields using the species finder, which will be semicolon-separated on the parent query form. The query results will include records that have one of the entered species for the corresponding section. If a user selects a more general category of a species, such as Vaccinia Virus, results will be returned with the species 'Vaccinia Virus' as well as the sub-categories of that species, such as Vaccinia virus (strain Ankara) [126794], and Vaccinia virus (strain WR) [10254]. If a user selects a specific species, such as Vaccinia virus (strain WR) [10254], only the results with the species 'Vaccinia virus (strain WR) [10254] ' will be returned.

1.1.5 Analysis Tools

All users will be able to utilize analysis tools after an Advanced Query, Simple Query, or Search has been executed. From the View Results screen (for advanced queries) or an Assay List page (for simple queries and searches), the user will have the ability to perform various types of analysis on results. The analysis tools are described in greater detail in Section 1.2.1.1

1.1.6 Browse Records by Allele

In addition to being able to search for records using the Advanced Query, Simple Query, or Search, all users can find records associated with a specific MHC allele by browsing records by allele. To browse records by allele, the user accesses the **Browse Records by Allele Name** page via the Query pull-down menu or the Quick Links on the IEDB home page. A table will be displayed with all the MHC alleles contained in the IEDB. The number of records associated

with each allele will be shown as a link. Clicking on a link will display the records associated with the selected allele.

1.1.7 Browse Records by Source Species

In addition to being able to search for records using the Advanced Query, Simple Query, or Search, all users can find records associated with a specific epitope source species by browsing records by species. To browse records by species, the user accesses the **Browse Records by Source Species** page via the Query pull-down menu or the Quick Links on the IEDB home page. A table will be displayed with all the epitope source species contained in the IEDB. The number of records associated with each species will be shown as a link. Clicking on a link will display the records associated with the selected species.

1.2 Resources

There are five major types of resource features available to users. Users can utilize the analysis and epitope prediction capabilities of the Analysis Resource. They can also make use of the IEDB forum to discuss and exchange information and ideas. In addition, users can export the contents of the database for their own use and manipulation, and download an ever increasing number of documents from the website. Finally, links to other websites of interest and relevance are provided.

1.2.1 Analysis Resource

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 will make users aware of available solutions to their problems. The tools will be evaluated and documented on the IEDB website, making it easier for users to choose the appropriate tool for a given task. All of the information contained within the Analysis Resource, including analysis tools and algorithms developed by the IEDB staff, will be freely available to the scientific community.

A major community outreach and feedback mechanism for the Analysis Resource will be the Analytical Tool Workshop being organized by NIAID for November 4, 2005 in Bethesda, Maryland. In this forum, the features of the first release and our plans for future enhancements will be presented to a gathering of tool developers and users. Participants will be able to try out the tools and provide feedback. The meeting will also help to identify knowledge gaps for future research and development of improved T cell and antibody epitope analysis and prediction tools.

This document describes the tools that have been developed by the IEDB team for the first release of the Analysis Resource, thus giving the reader background material for the tool demonstrations. No claim is made that these are the best tools available, rather future evaluations will provide guidance to what tools should be removed and which new ones should be added.

The tools provided in the analysis resource fall into two categories – analysis tools and predictive tools. Analytical tools help extract and interpret data contained in the database. Currently four tool candidates exist - population coverage (Section 1.2.1.1.1), conservancy analysis (Section

1.2.1.1.2), epitope visualization (Section 1.2.1.1.3), and mapping B-cell epitopes to the PDB structures (Section 1.2.1.1.4).

Predictive 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. The main difference between analytical and predictive tools is that it is a nontrivial task to determine if the outputs of the latter are correct.

For predictive 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. One benefit of the IEDB is that it allows implementing methods to automatically generate new predictive tools as the database grows

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

- Proteasomal cleavage

- TAP transport

- Binding to MHC class I / II

- B-cell epitope prediction

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

1.2.1.1 Analysis Tools

1.2.1.1.1 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 is required and a suitable algorithm is needed for this specific purpose (Bui et al. Bioinformatics 2005, submitted). 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 x 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 combination, 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.

We have compiled HLA allele genotypic frequencies from dbMHC database (http://www.ncbi.nlm.nih.gov/mhc/). At present, dbMHC provides allele frequencies for 78 populations and 11 different geographical areas. It is envisioned that the compiled data will be updated regularly as further HLA frequency data become available. Furthermore, customized frequency data can be utilized in the calculation should studies of specific and particular patient populations be of interest to a given user. Multiple population coverages can be simultaneously calculated and an average population coverage is generated. Since MHC class I and II restricted epitopes elicit immune responses from two different T cell populations (CTL and HTL, respectively), the program provides three different coverage calculation modes – (1) class I separate, (2) class II separate, and (3) class I and class II combined.

1.2.1.1.2   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.

1.2.1.1.3   Visualization Tool

The convenient and easy to use EpitopeViewer has been developed for three-dimensional visualization of immune epitopes and analyses of their interactions with antigen-specific

receptors of the immune system (antibodies and/or T cell receptors, MHC) for structures available in the Protein Data Bank (PDB). The EpitopeViewer is based on the Molecular Biology Toolkit (MBT; http://mbt.sdsc.edu/) developed at the San Diego Supercomputer Center (SDSC). It uses data both from the PDB and the IEDB, and visualizes one epitope (one PDB structure) at a time.

The EpitopeViewer provides the following functionality:

- Link to the PDB web-page displaying a particular structure.

- Visualization of the 3D structure of epitope/antigen and/or its complex with immune receptor(s) as curated within the IEDB and available in the PDB.

- Visualization of sequences of epitope/antigen and immune receptor(s).

- 3D-visualization of intermolecular (epitope-paratope, epitope-antibody CDR, epitope-MHC, pMHC-TCR, pMHC-TCR CDR), inter-atom and inter-residue interactions curated within the IEDB and/or calculated on the fly from the PDB file with essential details (contact type, atoms, distance) provided;

- 2D-plot of inter-residue interactions between epitope and immune receptor.

- Generation of publication-quality pictures of structures, sequences, and plots of contacting residues.

The EpitopeViewer (developed as a Java-application) runs via Java Webstart with capability of fully automated setup of Java3D on the client machine. The tool has been tested on a hundred epitopes of various types whose structural data are available in the PDB, including B-cell epitopes of different chemical types (peptides, globular proteins, carbohydrates, DNAs, and haptens), as well as T-cell epitopes in complexes with MHC class I and II, and ternary complexes TCR-pMHC.

1.2.1.1.4 Tool for mapping linear epitopes to the PDB structures

EpitopeMappingToPDB is an automated tool that maps linear epitopes to three-dimensional protein structures within the PDB by sequence similarity of antigen and proteins in the PDB. For input, the tool uses the SwissProt ID of the antigen protein, the epitope sequence, and the position of the epitope in the antigen sequence as curated within the IEDB or input by the user. The tool applies the NCBI BLAST algorithm for performing sequence homology search, and provides options for the sophisticated user to choose cutoff values on parameters used in the search programs (such as e-value and penalty on gap initiation and gap extension). The tool output page displays the alignment between the query sequence of the antigen containing the epitope and the sequence from the PDB representing significant hits (matches). The region within the epitope is highlighted in the alignment, and the sequence identity for the epitope and homologous region is provided. The EpitopeViewer application for visualization of homologous epitope/antigen and its further structural analysis is launched from the output page.

1.2.1.2 Predictive tools

1.2.1.2.1 MHC binding prediction methods

Incorporating a prediction method means implementing a computer algorithm that automatically generates updated prediction tools as new data become available in the IEDB. While it is imaginable to implement methods for all kinds of predictions, we decided to start with the best defined field, namely peptide binding to MHC molecules. Here, implementing methods will allow extending predictions to new alleles and peptides of different lengths as more data become available as well as continuously making predictions more accurate.

Implementing methods in the IEDB requires a tight integration of the method algorithm with the database. To have working examples for the first release of the analysis resource, the methods proposed below are those best known to members of the IEDB team. They are therefore a practical first choice, which can later be revised as the evaluation of different methods may show others to be superior.

1.2.1.2.1.1 Artificial neural network

Artificial neural networks 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. The training data in our case are peptide sequences with quantitative affinities for a specific MHC molecule.

Many different implementations of artificial neural networks exist. The one utilized here is described for HLA-A2 binding predictions by Nielsen et al. (Protein Science, 2003) and has recently been applied to a number of different alleles (http://www.cbs.dtu.dk/services/NetMHC/). The program used for training the networks was developed by Søren Brunak in Fortran77, and has been made to run on a number of platforms. A separate program developed in C by Morten Nielsen is used by the web server to make predictions.

1.2.1.2.1.2 ARB – matrix predictions

Average relative binding (ARB) matrix binding prediction method is based on the assumption that each residue along the peptide molecule independently contributes to binding affinity. When a residue R occurs at position i in the peptide, it is assumed to contribute a constant amount of R_i to the free energy of binding of the peptide. The effect of each of the 20 possible amino acids at each possible position along the peptide sequence, therefore, can be estimated by a matrix of coefficients. The overall binding propensity of each peptide sequence, an algorithm "score", is calculated by multiplying the R_i coefficients. Predict IC_{50} values which provide quantitative K_D (IC_{50}) predictions are then calculated by mathematical transformations of the algorithm scores (Bui et al., Immunogenetics 2005).

1.2.1.2.1.3 SMM – matrix predictions

The SMM method described in (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. The SMM algorithm is currently implemented in C++ code as a stand-alone application.

1.2.1.2.2   Proteasomal cleavage and TAP transport predictions

For the prediction of MHC class I antigen processing, we will incorporate 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. From this experimental data, SMM matrices (Peters and Sette, BMC Bioinformatics, 2005) have been generated. The proteasomal cleavage predictions evaluate how efficient a peptide – or its N-terminally prolonged precursors - can be liberated from its source protein. The TAP transport predictions evaluate how efficient a peptide or its N-terminal prolonged precursors are transported into the ER by TAP (Peters, J Immunol, 2003). Taken together and combined with MHC class I binding predictions, this allows to predict the efficiency with which a peptide is presented on the cell surface.

1.2.1.2.3   Sequence based antibody epitope prediction

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 for implementation based on their popularity and coverage of different categories.

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

In general, antibody epitope prediction tools exist only as described in published literatures or stand-alone applications. Therefore, to make these tools available and conveniently accessible on a website, the IEDB team has implemented them as a Java servlet application. Several computer algorithms were developed to perform different calculation types based on the published procedures. The implemented tools were validated based on the specific input and output data if available from the publication that describes the tool.

1.2.1.3 References

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 M, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. Bioinformatics (submitted)

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

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, Holzhutter HG. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. J Immunol. 2003 Aug 15;171(4):1741-9.

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.

1.2.2 Forums

All users can view IEDB forums. New forums will be created by an Application Administrator in response to feedback, help requests, and frequently asked questions. All of the discussion topics within a forum are displayed when viewing a forum. Help content for the forums is provided by Jive Forums. All users can search for discussion topics in forums, but only registered users can submit and reply to a discussion topic in a forum.

1.2.3 Database Export

All users are able to download export files. An export file will contain a pre-established subset of data from the IEDB. The files will be available for downloading in XML format. The XML file format allows data to be presented in a hierarchical manner. There are many tools available for processing and displaying this type of data. The available export files will be updated on a periodic basis.

1.2.4 Downloads

All users are able to download documents. Documents include files such as Introduction to IEDB and Analysis Resource, IEDB Ontology, Curation Instructions, IEDB XSD Documentation, and other miscellaneous files that do not contain research data. The file format will vary depending on the file.

1.2.5 View External Links

All users are able to view external links related to the IEDB and use the links if they choose. The available links are broken down into the following categories:

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

1.3 Account Information

1.3.1 Become a Registered User

IEDB has two types of users: anonymous users and registered users. Anonymous users are persons who use the IEDB application and do not register with IEDB. Users can become registered with the IEDB by submitting an e-mail address and a password, and indicating the institution they belong to, the country where they live, and if they would like a newsletter subscription. Registered users are able to use features that are unavailable to anonymous users.

1.3.2 Registered User Login

Users must be registered with IEDB in order to log in to the system. Users will be automatically prompted to log in to the IEDB upon attempting to use a function that requires user registration.

1.3.3 Modify Account Information

Account modifications consist of e-mail address, institution, country, password, and/or news letter subscription changes. Users must already be registered in order to have an account to modify.

1.4 Support Overview

The IEDB offers numerous ways to learn how to use the IEDB and to receive help if a problem is encountered. On almost every page of the IEDB, instructions are listed in a box at the top of the screen that describes how to perform the most important functions on the page. The instructions can be removed from all pages by clicking on the Hide Instructions link in the Quick Links box on the home page. Additionally, a more detailed description of each page can be accessed by clicking the 🕐 button to the right of a page title. When performing an Advanced Query, a hint for each field can be seen by moving the cursor over a field name. The hint will define the field, describe how to enter a value, and give an example. Click on a field name for a more detailed description of the field. The online help can also be accessed by moving the cursor over the Support tab and selecting Online Help. The online help describes all features of the IEDB in detail.

Numerous tours are also available which describe how to perform different types of queries and provide an overview of the IEDB. Additionally, forums are available so users can view, post, and reply to discussion topics relating to the IEDB or immunology in general.

The IEDB has done its best to make the application intuitive and to describe how to use all features of the IEDB. However, questions may arise that the online help, tours, and forums do not answer. In these cases, users can submit feedback or submit a help request.

1.4.1 View Online Help

All users are able to view the online help. Online help is designed to be an online user guide. It describes in detail all of the features that the system offers, explains how users go about doing everything the system has to offer, and tries to answer any questions that users might have when using the system.

1.4.2 Hide or Show Instructions

Each page of the IEDB application, other than the home page, has instructions under the page title explaining certain features on the page. On the IEDB home page, when the application is opened there is a Hide Instructions link in the Quick Links box. Click on this link to remove the instructions listed on every page. After the Hide Instructions link has been clicked, a Show Instructions link will be displayed. Click on the Show Instructions link to have the instructions displayed again on every page.

1.4.3 Submit Feedback

All users are able to submit feedback and let the IEDB team know what they think about the system. Users can express what they like or dislike about the system, as well as additional 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.

1.4.4 Submit Help Request

Only users logged in to the system can submit help requests. Help requests will be sent to the help desk. Help requests should consist of problems that users have with the application, such as a certain function of the system not working, and not questions on how to use features of the application.   The online help is designed to answer questions regarding features of the application. A full name, day phone number, and description of the problem are required to submit a help request. After submitting a help request, a confirmation e-mail will be sent to the user's registered 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.

1.4.5 Tour the IEDB

Several tours are available that describe the IEDB in general, how to perform different types of queries, and how to interpret query results. The tours are available in Flash Media format and are divided into sections. The length, description, and script, which can be downloaded, are listed for each tour on the **Tour the IEDB** page. The tours and their descriptions are listed in Table 1.2

Tour Category	Tour Name	Description
Overview	IEDB Overview	Provides an overview of the Immune Epitope Database and Analysis Resource or IEDB. The following modules are discussed: Query, Analysis tools, Discussion forum, and Online help.
Query Development Overviews	Search	Describes how to perform a Search.
	Simple Query	The Simple Query input form and how to input search criteria are discussed in this tour.
	Advanced Query	Outlines search criteria development and how registered users save advanced queries and load saved queries.
Query Results Overviews	Search and Simple Query	The result summary page and how to analyze and download results are discussed for a Search or Simple Query in this tour.
	Advanced Query	Describes how to format, download, and analyze Advanced Query results.
Sample Queries	Advanced Epitope Structure Query	Uses regular expression syntax to retrieve epitopes with a specific linear sequence using an Advanced Query. In this query, a regular expression is entered into the Epitope Linear Sequence field.
	Advanced Epitope Structure and Source Query 1	Demonstrates how to retrieve epitopes with a specific linear sequence and source name using an Advanced Query. The search fields include Epitope Linear Sequence and Epitope Source Name.
	Advanced Epitope Structure and Source Query 2	Demonstrates how to retrieve peptide epitopes for a species and use the custom report to format query results. The search fields include Epitope Chemical Type and Epitope Source Species.
	Advanced B Cell Query	Demonstrates how to find all epitopes with B Cell Response assays using monoclonal antibodies that react to a particular protein from a specific species strain using an Advanced Query. The search fields include B Cell Antibody Type and Epitope Source Name.
	Simple B Cell Query	Demonstrates how to find all epitopes with B Cell Response assays using monoclonal antibodies that react to a particular protein from a specific species strain using a Simple Query. The search fields include B Cell Antibody Type and Epitope Protein or Gene.
	Simple T Cell Query	Demonstrates how to retrieve T cell epitopes with a given MHC restriction using a Simple Query. The Assay Categories and MHC Restriction fields are used.

Table 1.2 Flash media tours describing the basic IEDB functionality are available

2   Antibody and T Cell Epitopes

The following table lists the number of distinct T cell and antibody epitopes by source species that were in the IEDB as of 31 December 2005. The source species is the species from which the epitopes originate, NOT the organism that one might be immune to necessarily. For this reason, the list contains not only Category A-C pathogens, but also mouse epitopes that were generated from mice. The NCBI taxonomy ID is provided for each source species and strain. The sum of antibody and T cell epitopes may exceed the listed subtotal if an epitope is both a T cell epitope and an antibody epitope, such as Dengue virus type 2 in row 50. This list differs from the one available on the **Browse Records by Source Species** web page because the table below only counts distinct epitopes, not records. A **distinct epitope** is an epitope with given characteristics regardless of the reference and has a positive qualitative binding value as measured by assay. The **Browse Records by Source Species** list provides a count of records, which may include molecular structures in the database that have negative qualitative binding values. This comes about because the IEDB collects both positive and negative data, a helpful feature for epitope prediction tool developers. The reader can obtain more information about the epitopes of a specific source species by going to http://www.immuneepitope.org and performing a query.

Row	Taxon. ID	Species/Strain Name	Antibody	T Cell	Subtotal
1	228399	Actinobacillus pleuropneumoniae serovar 1 str. 4074	2	0	2
2	1049	Allochromatium vinosum	0	1	1
3	5599	Alternaria alternata	3	0	3
4	4212	Ambrosia artemisiifolia	0	1	1
5	171929	Anacardium occidentale	11	0	11
6	770	Anaplasma marginale	14	6	20
7	234826	Anaplasma marginale str. St. Maries	0	1	1
8	46607	Andes virus	50	0	50
9	3818	Arachis hypogaea	39	8	47
10	5085	Aspergillus fumigatus	19	0	19
11	5064	Aspergillus restrictus	13	0	13
12	7604	Asterias rubens	0	1	1
13	31563	Avian rotavirus A	4	0	4
14	5865	Babesia bovis	0	10	10
15	5868	Babesia microti	2	0	2
16	1392	Bacillus anthracis	2	5	7
17	37962	Bayou virus	1	0	1
18	520	Bordetella pertussis	0	3	3
19	12455	Borna disease virus	0	1	1
20	29518	Borrelia afzelii	0	1	1
21	139	Borrelia burgdorferi	0	13	13
22	9913	Bos taurus	19	31	50
23	11303	Bovine ephemeral fever virus	2	0	2
24	11099	Bovine viral diarrhea virus 1	1	0	1
25	3707	Brassica juncea	9	0	9
26	29459	Brucella melitensis	1	0	1

Row	Taxon. ID	Species/Strain Name	Antibody	T Cell	Subtotal
27	224914	Brucella melitensis 16M	1	0	1
28	236	Brucella melitensis biovar Ovis	2	0	2
29	6239	Caenorhabditis elegans	0	1	1
30	5476	Candida albicans	2	15	17
31	74724	Canine calicivirus	2	0	2
32	9534	Cercopithecus aethiops	0	1	1
33	7155	Chironomus thummi thummi	3	0	3
34	810	Chlamydia	0	1	1
35	813	Chlamydia trachomatis	10	8	18
36	83558	Chlamydophila pneumoniae	0	8	8
37	182082	Chlamydophila pneumoniae TW-183	0	9	9
38	11096	Classical swine fever virus	1	0	1
39	1491	Clostridium botulinum	0	6	6
40	36826	Clostridium botulinum A	44	0	44
41	1502	Clostridium perfringens	1	0	1
42	1513	Clostridium tetani	0	22	22
43	8932	Columba livia	0	7	7
44	1717	Corynebacterium diphtheriae	0	1	1
45	10034	Cricetus cricetus	0	1	1
46	3369	Cryptomeria japonica	4	1	5
47	10358	Cytomegalovirus	0	23	23
48	11052	Dengue virus group	1	0	1
49	11053	Dengue virus type 1	1	0	1
50	11060	Dengue virus type 2	8	2	9
51	11065	Dengue virus type 2 (NGC-prototype)	10	8	18
52	31635	Dengue virus type 2 (strain 16681-PDK53)	1	2	3
53	11064	Dengue virus type 2 (strain Jamaica)	4	0	4
54	11066	Dengue virus type 2 (strain PR159/S1)	2	2	2
55	11068	Dengue virus type 2 (strain PUO-218)	1	1	1
56	11069	Dengue virus type 3	0	9	9
57	11070	Dengue virus type 4	0	10	10
58	6954	Dermatophagoides farinae	0	3	3
59	6956	Dermatophagoides pteronyssinus	0	6	6
60	7227	Drosophila melanogaster	0	1	1
61	11268	Ebola virus	0	1	1
62	129003	Ebola virus strain Reston	2	0	2
63	128952	Ebola virus strain Zaire Mayinga	9	3	12
64	31705	Echovirus 11 (strain Gregory)	0	1	1
65	12104	Encephalomyocarditis virus	0	0	0
66	5759	Entamoeba histolytica	4	0	4
67	10710	Enterobacteria phage lambda	0	7	7
68	11665	Equine infectious anemia virus	0	1	1
69	11671	Equine infectious anemia virus (STRAIN WSU5)	0	6	6
70	562	Escherichia coli	0	3	3
71	83333	Escherichia coli K12	0	7	7
72	83334	Escherichia coli O157:H7	1	0	1

Row	Taxon. ID	Species/Strain Name	Antibody	T Cell	Subtotal
73	11978	Feline calicivirus	4	0	4
74	9685	Felis catus	3	2	5
75	12110	Foot-and-mouth disease virus	1	0	1
76	73482	Foot-and-mouth disease virus (strain O1)	1	0	1
77	12118	Foot-and-mouth disease virus - type O	1	0	1
78	263	Francisella tularensis	0	3	3
79	11795	Friend murine leukemia virus	0	2	2
80	11797	Friend murine leukemia virus (ISOLATE FB29)	0	1	1
81	9031	Gallus gallus	9	22	31
82	3847	Glycine max	17	0	17
83	727	Haemophilus influenzae	0	1	1
84	11602	Hantaan virus 76-118	3	2	5
85	210	Helicobacter pylori	1	0	1
86	12092	Hepatitis A virus	16	0	16
87	12098	Hepatitis A virus (STRAIN HM-175)	82	0	82
88	10407	Hepatitis B virus	1	67	67
89	10411	Hepatitis B virus (STRAIN ALPHA1)	0	10	10
90	10414	Hepatitis B virus (STRAIN LSH / CHIMPANZEE ISOLATE)	0	1	1
91	10409	Hepatitis B virus (SUBTYPE ADR4)	0	10	10
92	10410	Hepatitis B virus (SUBTYPE ADW / STRAIN 991)	0	4	4
93	10412	Hepatitis B virus (SUBTYPE ADW / STRAIN INDONESIA/PIDW420)	0	1	1
94	10413	Hepatitis B virus (SUBTYPE ADW / STRAIN JAPAN/PJDW233)	0	20	20
95	10415	Hepatitis B virus (SUBTYPE ADW / STRAIN OKINAWA/PODW282)	0	10	10
96	10408	Hepatitis B virus (SUBTYPE ADW2)	0	33	33
97	31512	Hepatitis B virus (subtype ADR / mutant)	0	1	1
98	31515	Hepatitis B virus (subtype ADW2 variant SF)	0	2	2
99	106821	Hepatitis B virus subtype ADW	0	2	2
100	10419	Hepatitis B virus subtype ADYW	0	6	6
101	10418	Hepatitis B virus subtype AYW	0	68	68
102	11103	Hepatitis C virus	0	9	9
103	11104	Hepatitis C virus (isolate 1)	0	20	20
104	11105	Hepatitis C virus (isolate BK)	0	39	39
105	11108	Hepatitis C virus (isolate H)	0	17	17
106	63746	Hepatitis C virus (isolate H77)	1	0	1
107	11111	Hepatitis C virus (isolate HC-J2)	0	1	1
108	11112	Hepatitis C virus (isolate HC-J5)	0	1	1
109	11113	Hepatitis C virus (isolate HC-J6)	0	1	1
110	11115	Hepatitis C virus (isolate HC-J8)	0	3	3
111	11110	Hepatitis C virus (isolate HCT18)	0	1	1
112	11116	Hepatitis C virus (isolate Japanese)	0	21	21
113	11117	Hepatitis C virus (isolate TH)	0	3	3
114	31645	Hepatitis C virus (isolate Taiwan)	0	17	17
115	41856	Hepatitis C virus genotype 1	0	18	18

Row	Taxon. ID	Species/Strain Name	Antibody	T Cell	Subtotal
116	31647	Hepatitis C virus subtype 1b	1	11	11
117	45410	Hepatitis b virus (subtype ADW4 / strain brazil / isolate w4b)	0	2	2
118	12475	Hepatitis delta virus	0	5	5
119	10426	Hepatitis delta virus (ISOLATE NAURU)	0	3	3
120	31764	Hepatitis delta virus (isolate woodchuck)	0	1	1
121	10299	Herpes simplex virus (type 1 / strain 17)	0	0	0
122	10301	Herpes simplex virus (type 1 / strain Angelotti)	0	2	2
123	3981	Hevea brasiliensis	34	0	34
124	9606	Homo sapiens	27	1388	1413
125	11927	Human T-cell lymphotrophic virus type 1 (Caribbean isolate)	0	10	10
126	11926	Human T-cell lymphotrophic virus type 1 (strain ATK)	0	2	2
127	11908	Human T-lymphotropic virus 1	1	6	7
128	10515	Human adenovirus type 2	0	1	1
129	28284	Human adenovirus type 40	0	1	1
130	28285	Human adenovirus type 5	0	1	1
131	10519	Human adenovirus type 7	0	1	1
132	11137	Human coronavirus 229E	0	2	2
133	10298	Human herpesvirus 1	0	1	1
134	10306	Human herpesvirus 1 strain KOS	0	2	2
135	10310	Human herpesvirus 2	0	6	6
136	10313	Human herpesvirus 2 strain 333	0	1	1
137	10315	Human herpesvirus 2 strain HG52	0	4	4
138	10376	Human herpesvirus 4	0	10	10
139	10377	Human herpesvirus 4 (strain B95-8)	0	60	60
140	10359	Human herpesvirus 5	0	6	6
141	10360	Human herpesvirus 5 strain AD169	0	6	6
142	10363	Human herpesvirus 5 strain Towne	0	4	4
143	10370	Human herpesvirus 6 (strain Uganda-1102)	0	1	1
144	57278	Human herpesvirus 7 strain JI	0	1	1
145	12721	Human immunodeficiency virus	0	9	9
146	11676	Human immunodeficiency virus 1	0	15	15
147	11685	Human immunodeficiency virus type 1 (ARV2/SF2 ISOLATE)	0	6	6
148	11678	Human immunodeficiency virus type 1 (BH10 ISOLATE)	0	5	5
149	11684	Human immunodeficiency virus type 1 (BH8 ISOLATE)	0	1	1
150	11693	Human immunodeficiency virus type 1 (BRAIN ISOLATE)	0	2	2
151	11686	Human immunodeficiency virus type 1 (BRU ISOLATE)	0	3	3
152	11679	Human immunodeficiency virus type 1 (CLONE 12)	0	2	2
153	11689	Human immunodeficiency virus type 1 (ELI ISOLATE)	0	8	8
154	11706	Human immunodeficiency virus type 1 (HXB2 ISOLATE)	0	28	28

Row	Taxon. ID	Species/Strain Name	Antibody	T Cell	Subtotal
155	11707	Human immunodeficiency virus type 1 (HXB3 ISOLATE)	0	4	4
156	11688	Human immunodeficiency virus type 1 (JRCSF ISOLATE)	0	33	33
157	11697	Human immunodeficiency virus type 1 (MAL ISOLATE)	0	10	10
158	11704	Human immunodeficiency virus type 1 (MFA ISOLATE)	0	2	2
159	11696	Human immunodeficiency virus type 1 (MN ISOLATE)	0	19	19
160	11695	Human immunodeficiency virus type 1 (NDK ISOLATE)	0	3	3
161	11698	Human immunodeficiency virus type 1 (NEW YORK-5 ISOLATE)	0	1	1
162	11699	Human immunodeficiency virus type 1 (OYI ISOLATE)	0	0	0
163	11700	Human immunodeficiency virus type 1 (PV22 ISOLATE)	0	3	3
164	11701	Human immunodeficiency virus type 1 (RF/HAT ISOLATE)	0	24	24
165	11703	Human immunodeficiency virus type 1 (STRAIN UGANDAN / ISOLATE U455)	0	1	1
166	31678	Human immunodeficiency virus type 1 (WMJ1 isolate)	0	15	15
167	11705	Human immunodeficiency virus type 1 (WMJ2 ISOLATE)	0	1	1
168	11683	Human immunodeficiency virus type 1 (Z2/CDC-Z34 ISOLATE)	0	1	1
169	11708	Human immunodeficiency virus type 1 (ZAIRE 6 ISOLATE)	0	2	2
170	82834	Human immunodeficiency virus type 1 lw12.3 isolate	0	16	16
171	11717	Human immunodeficiency virus type 2 (ISOLATE GHANA-1)	0	1	1
172	11719	Human immunodeficiency virus type 2 (ISOLATE NIH-Z)	0	1	1
173	11721	Human immunodeficiency virus type 2 (ISOLATE ST)	0	1	1
174	10566	Human papillomavirus	0	2	2
175	10580	Human papillomavirus type 11	0	3	3
176	333760	Human papillomavirus type 16	1	27	28
177	333761	Human papillomavirus type 18	0	4	4
178	10585	Human papillomavirus type 31	0	0	0
179	10618	Human papillomavirus type 52	0	1	1
180	37115	Human papillomavirus type 59	0	1	1
181	10600	Human papillomavirus type 6b	0	5	5
182	10602	Human papillomavirus type me180	0	1	1
183	11260	Human respiratory syncytial virus A strain Long	0	3	3
184	11259	Human respiratory syncytial virus A2	0	1	1
185	10960	Human rotavirus (SEROTYPE 4 / STRAIN ST. THOMAS 3)	0	0	0

Row	Taxon. ID	Species/Strain Name	Antibody	T Cell	Subtotal
186	11320	Influenza A virus	7	37	43
187	135322	Influenza A virus (A/Ann Arbor/6/60(H2N2))	0	2	2
188	11325	Influenza A virus (A/Bangkok/1/79(H3N2))	0	11	11
189	135658	Influenza A virus (A/Environment/Hong Kong/437-4/99 (H5N1))	0	0	0
190	108859	Influenza A virus (A/Hong Kong/1/68(H3N2))	0	2	2
191	220500	Influenza A virus (A/Japan/305/57(H2N2))	1	18	18
192	11428	Influenza A virus (A/Leningrad/54/1(H1N1))	0	1	1
193	107471	Influenza A virus (A/Louisiana/1/93(H3N2))	0	0	0
194	252233	Influenza A virus (A/Memphis/102/72(H3N2))	0	3	3
195	228928	Influenza A virus (A/Memphis/31/98(H3N2))	4	0	4
196	260805	Influenza A virus (A/NT/60/68/(H3N2))	0	14	14
197	132734	Influenza A virus (A/Netherlands/018/94 (H3N2))	0	0	0
198	132842	Influenza A virus (A/Netherlands/889/91 (H3N2))	0	1	1
199	223935	Influenza A virus (A/Okuda/57(H2N2))	0	9	9
200	119209	Influenza A virus (A/Philippines/2/82(H3N2))	1	0	1
201	83922	Influenza A virus (A/Port Chalmers/1/73(H3N2))	0	1	1
202	107478	Influenza A virus (A/Puerto Rico/1/97(H3N2))	1	0	1
203	211044	Influenza A virus (A/Puerto Rico/8/34(H1N1))	3	171	174
204	183764	Influenza A virus (A/Puerto Rico/8/34/Mount Sinai(H1N1))	1	8	8
205	183796	Influenza A virus (A/Texas/1/77(H3N2))	1	8	8
206	225083	Influenza A virus (A/USSR/90/77(H1N1))	7	3	10
207	11480	Influenza A virus (A/Udorn/307/72(H3N2))	0	3	3
208	62522	Influenza A virus (A/Victoria/1/93(H3N2))	0	1	1
209	78586	Influenza A virus (A/Victoria/3/75(H3N2))	0	5	5
210	145855	Influenza A virus (A/WSN/33(H1N1))	2	0	2
211	63106	Influenza A virus (A/Wuhan/359/95(H3N2))	3	0	3
212	132504	Influenza A virus (A/X-31(H3N2))	2	4	6
213	36418	Influenza A virus (A/chicken/Brescia/1902(H7N7))	0	1	1
214	11363	Influenza A virus (A/duck/Hokkaido/10/85(H3))	0	1	1
215	258040	Influenza A virus (A/duck/Hokkaido/107/01(H2N3))	0	1	1
216	95888	Influenza A virus (A/mallard duck/PA/10218/84(H5N2))	1	0	1
217	169169	Influenza A virus (A/swine/Cotes d'Armor/1482/99(H1N1))	0	1	1
218	219637	Influenza A virus (A/swine/Hong Kong/6/76(H3N2))	0	3	3
219	219641	Influenza A virus (A/swine/Hong Kong/81/78(H3N2))	0	2	2
220	145307	Influenza A virus (A/swine/Hong Kong/9/98(H9N2))	2	0	2
221	11504	Influenza A virus (A/swine/Indiana/1726/88(H1N1))	0	1	1
222	260809	Influenza A virus (A/swine/Iowa/15/30(H1N1))	0	2	2
223	11517	Influenza A virus (A/swine/Ukkel/1/84(H3N2))	0	1	1
224	88305	Influenza A virus (A/swine/Wisconsin/1/61(H1N1))	0	2	2
225	11357	Influenza A virus (STRAIN A/DUCK/HOKKAIDO/5/77)	0	1	1
226	41857	Influenza A virus H3N2	0	4	4

Row	Taxon. ID	Species/Strain Name	Antibody	T Cell	Subtotal
227	11520	Influenza B virus	1	0	1
228	184816	Influenza B virus (B/Kadoma/122/99)	2	0	2
229	107412	Influenza B virus (B/Lee/40)	0	2	2
230	107417	Influenza B virus (B/Singapore/222/79)	0	1	1
231	107418	Influenza B virus (B/Victoria/2/87)	0	0	0
232	11553	Influenza C virus (C/Ann Arbor/1/50)	1	0	1
233	197911	Influenzavirus A	0	12	12
234	42097	Isla Vista virus	1	0	1
235	11072	Japanese encephalitis virus	2	0	2
236	11075	Japanese encephalitis virus strain JAOARS982	4	2	6
237	11076	Japanese encephalitis virus strain Nakayama	2	0	2
238	51240	Juglans regia	1	0	1
239	11619	Junin virus	0	0	0
240	13101	Juniperus ashei	5	0	5
241	42894	Khabarovsk virus	1	0	1
242	10638	Kilham polyomavirus	0	0	0
243	571	Klebsiella oxytoca	0	0	0
244	573	Klebsiella pneumoniae	15	0	15
245	11078	Kunjin virus (STRAIN MRM61C)	0	3	3
246	33728	Lake Victoria marburg virus - Popp	1	0	1
247	33727	Lake Victoria marburgvirus - Musoke	1	0	1
248	11621	Lassa virus GA391	0	1	1
249	11622	Lassa virus Josiah	0	4	4
250	5664	Leishmania major	0	1	1
251	1639	Listeria monocytogenes	0	20	20
252	169963	Listeria monocytogenes EGD-e	5	34	39
253	4522	Lolium perenne	7	10	17
254	11623	Lymphocytic choriomeningitis virus	0	17	17
255	11624	Lymphocytic choriomeningitis virus (strain Armstrong)	0	29	29
256	11627	Lymphocytic choriomeningitis virus (strain WE)	0	5	5
257	11628	Machupo virus	0	0	0
258	7130	Manduca sexta	0	2	2
259	45201	Mannheimia haemolytica serotype 1	13	0	13
260	11237	Measles virus strain IP-3-CA	0	1	1
261	11239	Measles virus strain Yamagata-1	0	3	3
262	11629	Mopeia virus	0	1	1
263	11759	Mouse mammary tumor virus (STRAIN C3H)	0	0	0
264	11171	Mumps virus (STRAIN MIYAHARA VACCINE)	0	0	0
265	33708	Murid herpesvirus 4	0	3	3
266	10367	Murine cytomegalovirus (strain Smith)	0	2	2
267	76344	Murine hepatitis virus strain 2	0	0	0
268	12760	Murine hepatitis virus strain 4	0	2	2
269	11144	Murine hepatitis virus strain JHM	0	1	1
270	11079	Murray Valley encephalitis virus	0	3	3
271	10090	Mus musculus	4	80	83

Row	Taxon. ID	Species/Strain Name	Antibody	T Cell	Subtotal
272	10092	Mus musculus domesticus	0	12	12
273	1764	Mycobacterium avium	0	2	2
274	1765	Mycobacterium bovis	0	21	21
275	33892	Mycobacterium bovis BCG	7	37	40
276	1767	Mycobacterium intracellulare	1	0	1
277	1769	Mycobacterium leprae	22	51	72
278	272631	Mycobacterium leprae TN	0	3	3
279	1771	Mycobacterium phlei	0	1	1
280	1773	Mycobacterium tuberculosis	13	279	292
281	83331	Mycobacterium tuberculosis CDC1551	0	13	13
282	83332	Mycobacterium tuberculosis H37Rv	29	146	158
283	485	Neisseria gonorrhoeae	0	9	9
284	487	Neisseria meningitidis	0	1	1
285	122928	Norovirus genogroup 1	1	0	1
286	122929	Norovirus genogroup 2	1	0	1
287	42764	Oliveros virus	0	0	0
288	784	Orientia tsutsugamushi	3	0	3
289	9986	Oryctolagus cuniculus	0	3	3
290	9940	Ovis aries	0	1	1
291	1406	Paenibacillus polymyxa	0	1	1
292	9598	Pan troglodytes	0	13	13
293	5076	Penicillium chrysogenum	10	0	10
294	6978	Periplaneta americana	3	0	3
295	15957	Phleum pratense	21	0	21
296	9755	Physeter catodon	0	25	25
297	11630	Pichinde virus	0	4	4
298	3888	Pisum sativum	0	1	1
299	5821	Plasmodium berghei	0	2	2
300	5833	Plasmodium falciparum	1	75	76
301	5835	Plasmodium falciparum (isolate CAMP / Malaysia)	0	1	1
302	5843	Plasmodium falciparum (isolate NF54)	0	6	6
303	57270	Plasmodium falciparum (isolate PALO ALTO / UGANDA)	0	5	5
304	5848	Plasmodium falciparum (isolate WELLCOME)	0	1	1
305	36329	Plasmodium falciparum 3D7	4	134	138
306	5854	Plasmodium reichenowi	0	1	1
307	5855	Plasmodium vivax	3	4	4
308	5861	Plasmodium yoelii	0	1	1
309	73239	Plasmodium yoelii yoelii	0	4	4
310	138953	Poliovirus	1	0	1
311	28344	Porcine respiratory and reproductive syndrome virus	1	0	1
312	11151	Porcine transmissible gastroenteritis coronavirus (STRAIN PURDUE)	1	0	1
313	837	Porphyromonas gingivalis	0	1	1
314	11603	Prospect Hill virus	1	0	1
315	287	Pseudomonas aeruginosa	3	0	3

Row	Taxon. ID	Species/Strain Name	Antibody	T Cell	Subtotal
316	11604	Puumala virus	1	1	1
317	39000	Puumala virus (strain k27)	0	6	6
318	39002	Puumala virus (strain sotkamo/v-2969/81)	131	0	131
319	11292	Rabies virus	2	2	4
320	11295	Rabies virus (strain ERA)	4	6	10
321	11296	Rabies virus (strain HEP-FLURY)	1	1	2
322	37132	Rabies virus (strain Ontario fox)	3	0	3
323	103929	Rabies virus (strain Pasteur / PV)	1	2	3
324	39005	Rabies virus (strain ontario skunk)	0	2	2
325	45418	Rabies virus (strain vnukovo-32)	0	1	1
326	10116	Rattus norvegicus	6	54	60
327	12814	Respiratory syncytial virus	0	7	7
328	160061	Ricinus	0	1	1
329	3988	Ricinus communis	3	3	6
330	272944	Rickettsia conorii str. Malish 7	0	5	5
331	11241	Rinderpest virus	3	0	3
332	46920	Rio Mamore virus	1	0	1
333	37207	Rio Segundo virus	1	0	1
334	11041	Rubella virus	1	0	1
335	11045	Rubella virus (STRAIN THERIEN)	0	1	1
336	227859	SARS coronavirus	46	136	182
337	228407	SARS coronavirus BJ01	23	19	42
338	228415	SARS coronavirus CUHK-W1	17	0	17
339	227984	SARS coronavirus Tor2	51	5	56
340	228330	SARS coronavirus Urbani	2	0	2
341	45709	Sabia virus	0	1	1
342	4932	Saccharomyces cerevisiae	0	3	3
343	10381	Saimiriine herpesvirus 2	0	0	0
344	601	Salmonella typhi	2	6	8
345	602	Salmonella typhimurium	0	2	2
346	6183	Schistosoma mansoni	0	1	1
347	4896	Schizosaccharomyces pombe	0	1	1
348	11033	Semliki forest virus	1	0	1
349	11191	Sendai virus	0	2	2
350	11198	Sendai virus (Z)	0	1	1
351	11194	Sendai virus (strain Enders)	0	1	1
352	11610	Seoul virus SR-11	1	0	1
353	4182	Sesamum indicum	11	0	11
354	623	Shigella flexneri	3	1	4
355	11723	Simian immunodeficiency virus	0	119	119
356	11735	Simian immunodeficiency virus (K6W ISOLATE)	0	50	50
357	11733	Simian immunodeficiency virus (MM142-83 ISOLATE)	0	0	0
358	11738	Simian immunodeficiency virus (PBJ/BC13 ISOLATE) (SOOTY MANGABEY)	0	0	0
359	31683	Simian immunodeficiency virus - stm	0	3	3

Row	Taxon. ID	Species/Strain Name	Antibody	T Cell	Subtotal
360	10633	Simian virus 40	0	4	4
361	57667	Simian-Human immunodeficiency virus	0	1	1
362	37705	Sin Nombre virus	1	0	1
363	6584	Spisula solidissima	0	3	3
364	1280	Staphylococcus aureus	17	4	21
365	1309	Streptococcus mutans	1	0	1
366	1314	Streptococcus pyogenes	4	8	12
367	301449	Streptococcus pyogenes serotype M5	0	1	1
368	1915	Streptomyces lincolnensis	0	1	1
369	31523	Suid herpesvirus 1 (strain Indiana-Funkhauser / Becker)	0	1	1
370	9823	Sus scrofa	1	4	5
371	11631	Tacaribe virus	0	0	0
372	12124	Theiler's encephalomyelitis virus	0	8	8
373	12127	Theiler's encephalomyelitis virus (STRAIN GDVII)	0	0	0
374	11084	Tick-borne encephalitis virus	12	0	12
375	83192	Topografov virus	1	0	1
376	7788	Torpedo marmorata	0	3	3
377	5811	Toxoplasma gondii	7	8	15
378	160	Treponema pallidum	0	0	0
379	9337	Trichosurus vulpecula	0	7	7
380	4565	Triticum aestivum	0	20	20
381	5693	Trypanosoma cruzi	1	19	20
382	37133	Tula virus	1	0	1
383	10245	Vaccinia virus	0	25	25
384	126794	Vaccinia virus Ankara	0	2	2
385	10254	Vaccinia virus WR	0	2	2
386	11276	Vesicular stomatitis virus	0	2	2
387	11284	Vesicular stomatitis virus (strain Orsay)	0	0	0
388	666	Vibrio cholerae	12	0	12
389	127906	Vibrio cholerae O1	8	0	8
390	686	Vibrio cholerae O1 biovar eltor	3	0	3
391	11288	Viral hemorrhagic septicemia virus (STRAIN 07-71)	1	0	1
392	3972	Viscum album	17	0	17
393	11082	West Nile virus	4	14	18
394	10436	Woodchuck hepatitis virus w64 (ISOLATE PWS23)	0	1	1
395	8355	Xenopus laevis	0	7	7
396	11090	Yellow fever virus (STRAIN 17D)	1	9	10
397	630	Yersinia enterocolitica	0	1	1
398	632	Yersinia pestis	5	28	28
399	38018	unidentified bacteriophage	0	9	9
400	11309	unidentified influenza virus	0	12	12
		Total	1150	4547	5653

3 Scientific Publications

The following IEDB scientific papers were published in 2005 by the IEDB project team:

- Sette A, Fleri W, Peters B, Sathiamurthy M, Bui HH, Wilson S: "A Roadmap for the Immunomics of Category A-C Pathogens", *Immunity*, 22:155-161, 2005

- 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 Biology*, 3:e91, 2005

- Bjoern Peters and Alessandro Sette, Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method, *BMC Bioinformatics* 2005, 6:132

- 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.

- 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 Design and Implementation of the Immune Epitope Data Base and Analysis Resource", *Immunogenetics*

- Muthuraman Sathiamurthy, Bjoern Peters, Huynh-Hoa Bui, John Sidney, John Mokili, Stephen S Wilson, Ward Fleri, Deborah L McGuinness, Philip E Bourne, and Alessandro Sette, "An ontology for immune epitopes: application to the design of a broad scope database of immune reactivities".

- Peters B, Bui HH, Sidney J, Weng Z, Loffredo JT, Watkins DI, Mothe BR, Sette A., "A computational resource for the prediction of peptide binding to Indian rhesus macaque MHC class I molecules", Vaccine. 2005 Aug 29