Published in final edited form as:

J Immunol Methods. 2011 November 30; 374(1-2): 62–69. doi:10.1016/j.jim.2010.10.010.

Applications for T-cell epitope queries and tools in the Immune Epitope Database and Analysis Resource

Yohan Kim^a, Alessandro Sette^a, and Bjoern Peters^a

^aLa Jolla Institute for Allergy & Immunology 9420 Athena Circle La Jolla, CA 92037, USA

Abstract

The Immune Epitope Database and Analysis Resource (IEDB, <http://www.iedb.org>) hosts a continuously growing set of immune epitope data curated from the literature, as well as data submitted directly by experimental scientists. In addition, the IEDB hosts a collection of prediction tools for both MHC class I and II restricted T-cell epitopes that are regularly updated. In this review, we provide an overview of T-cell epitope data and prediction tools provided by the IEDB. We then illustrate effective use of these resources to support experimental studies. We focus on two applications, namely identification of conserved epitopes in novel strains of a previously studied pathogen, and prediction of novel T-cell epitopes to facilitate vaccine design. We address common questions and concerns faced by users, and identify patterns of usage that have proven successful.

Keywords

epitope conservation; epitope predictions; vaccine design; Major Histocompatibility Complex

1 Introduction

The goal of the IEDB (Peters et al., 2005; Peters and Sette, 2007) is to catalog and organize information related to T- and B-cell epitopes, as well as to provide tools to predict novel epitopes and to analyze known epitopes to gain new information about them. The IEDB website has been updated continuously based on user feedback, resulting in fundamentally revised versions of both the database itself (Vita et al., 2010) and the tools hosted in the associated Analysis Resource (Zhang et al., 2008). The continued use of the database and tools has broadened our understanding of what applications are possible and refined the methodology to achieve optimal results.

While a similar suite of data and tools exist for B-cell/antibody epitopes, they are outside of the scope for the present article. Moreover, details of curation process used to capture data from the literature (Vita et al., 2008), development of a formal ontology to represent the immune epitope data (Sathiamurthy et al., 2005), all available prediction and analysis tools (Zhang et al., 2008), and recent updates to the database structure (Vita et al., 2010) are

© 2010 Elsevier B.V. All rights reserved.

Corresponding author: Bjoern Peters (bpeters@liai.org) Tel: 858-752-6914 Fax: 858-752-6987.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

described elsewhere. We here present a summary of the currently available resources in the IEDB that are related to T-cell epitopes.

Multiple applications can benefit from identifying T-cell epitopes, including the design of prophylactic vaccines (Kaeche et al., 2002; Sette and Fikes, 2003; Purcell et al., 2007), therapeutics (Bousquet et al., 1998; Banchereau and Palucka, 2005), diagnostics (Brock et al., 2004; Pai et al., 2008), reagents for research (Altman et al., 1996), and de-immunization of biological drugs (Bryson et al., 2010). Designing vaccines against infectious agents is the most frequent application that can benefit from knowing T-cell epitopes. Basically, for a vaccine to induce the creation of a memory T-cell population capable of recognizing a pathogen, the vaccine has to contain T-cell epitopes from that pathogen. This does not mean that the vaccine itself has to be comprised of individual epitopes, but minimally that antigens that harbor the epitopes are being included. In the case of therapeutics (sometimes also called therapeutic vaccines), the goal is to modulate an existing immune response that either needs to be enhanced (cancer, chronic infection), or reduced (autoimmune diseases, allergy). Again, designing such a therapeutic requires knowing the epitope targets of the T-cell responses that are intended to be modulated. For diagnostics, the goal is to identify if a patient has been in contact with certain antigens in order to help identify a patient's disease status. The use of epitopes as diagnostics has the potential to distinguish infections with related pathogens, which can be difficult when relying on whole organism or antigen preparations, which can often be cross-recognized. Finally, a major application of T-cell epitopes is their direct use as reagents for basic research, as they allow tracking specific T-cell populations, notably in the form of MHC tetramers.

As more protein based drugs are developed, minimizing immunogenicity of these drugs becomes another important application of T-cell epitope identification. The term 'de-immunization' has been coined to refer to a set of technologies that reduce immunogenicity by removing T-cell epitopes from protein therapeutics (Jones et al., 2009). The procedure entails first identifying epitopes and their anchor residues for HLA binding, applying point mutations to these anchor residues to neutral ones such that mutations do not interfere with folding and function of proteins, and finally carrying out validations to confirm absence of immunogenicity of the mutated proteins (Scott and De Groot, 2010). The general T-cell epitope identification tools in the IEDB can be applied to this problem, but currently no dedicated tools exist in the IEDB to e.g. suggest what mutations in a protein drug are recommended.

Following the overview of resources in the IEDB, we present examples of applications how the IEDB can be used to 1) discover conserved epitopes in a novel antigen, and 2) predict novel epitopes to facilitate vaccine design. Overall, this article provides examples on how the IEDB data and tools can be used effectively for practical experimental applications, and addresses commonly encountered issues.

2 Overview of T-cell epitope related resources available in the IEDB

2.1 Data contained in the IEDB

The IEDB provides a catalog of experimentally characterized T-cell epitopes, as well as data on Major Histocompatibility Complex (MHC) binding and MHC ligand elution experiments. These three types of experimental data differ in what they convey: While MHC binding data shows a molecular association between a ligand (typically a peptide) and an MHC molecule, MHC ligand elution data shows that the eluted ligand is not only capable of binding an MHC molecule, but can also be created by the cellular antigen processing machinery. T-cell epitope mapping experiments, on the other hand, test the ability to engage a T-cell receptor (TCR), which requires binding to an MHC molecule and activation of the

TCR. The IEDB represents the molecular structures tested for MHC binding, MHC ligand elution or T-cell recognition, and the details of the experimental contexts in which these molecules were tested. Only peptides that test positive in a T-cell recognition assays are truly 'T-cell epitopes', while those that are only shown to be positive in a binding assay should be referred to as 'MHC binders' and those that can be eluted from MHC on the cells surface are 'naturally processed ligands'. Epitopes recognized in humans, nonhuman primates, rodents, pigs, cats and all other tested host species are included in the IEDB. Moreover, both positive and negative experimental results are captured. Lastly, the scope of the database to date includes data relating to epitope derived from all infectious diseases, including NIAID Category A, B and C priority pathogens (<http://www3.niaid.nih.gov/topics/BiodefenseRelated/Biodefense/research/CatA.html>), NIAID Emerging and Reemerging infectious diseases (<http://www3.niaid.nih.gov/topics/emerging/>), allergens, and autoantigens involved in autoimmune disorders. In particular, HIV epitopes are explicitly excluded, which can instead be found in the Los Alamos HIV Molecular Immunology Database (<http://www.hiv.lanl.gov>) (Korber et al.).

Since the IEDB was initiated 4 years ago, the data from 158,067 T-cell, 198,877 MHC binding, and 2,007 MHC ligand elution assays have been collected through either manual curation or direct submissions. Curated data from the literature covers 99% of all publicly available journal articles on peptidic epitopes mapped in infectious agents (excluding HIV) and 97.7% of those mapped in allergens. In addition, the curation of epitopes related to autoimmunity is expected to be completed by the end of 2010. In terms of direct submissions, 194,862 assays were submitted by investigators to the IEDB that would otherwise be unavailable to the public. TABLE 1 provides an overview of the journal articles available in each of the main categories of peptidic epitopes included in the IEDB and their level of completion.

2.2 Query Interfaces

The query interface of the IEDB has been completely redesigned in early 2009 for the IEDB 2.0 release (Vita et al., 2010). For instance, more frequently used query features are prominently displayed in the left panel of the home page (FIGURE 1). The query interface is divided into epitope structure (to query by e.g. peptide sequence), epitope source (to query by the antigen or organism in which an epitope is contained) and immunological contexts (to query by the type of assay, host organism of the response, or MHC restriction). The user can for example search for any epitope contained in Influenza A viruses that is recognized in a T-cell assay by human effector cells and restricted by MHC class II molecule. The results of this query are presented in FIGURE 2. For users with the need to perform more detailed queries, an advanced query mode is provided through the 'Search' menu item, allowing one to search any field in the database. In addition, epitope entries can be 'browsed' using features such as MHC alleles or source organisms using NCBI taxonomy tree (See 'Browse' tab at the top).

2.3 T-cell epitope prediction tools

Complementing the experimentally determined immune epitope data contained in the database, the Analysis Resource component of the IEDB provides tools for predicting and analyzing epitope data (<http://tools.immuneepitope.org>). Table 2 gives an overview of different T-cell relevant categories of tools available. The analysis tools provide the ability to map epitopes onto 3D protein structure of the source antigen or its homologue (Beaver et al., 2007), cluster epitopes into groups with similar sequences, calculate the degree of conservation of epitopes in a given set of proteins (Bui et al., 2007), or predict the

population coverage of a set of epitopes based on the frequency of the HLA alleles they are restricted by (Bui et al., 2006).

The T-cell epitope prediction tools are divided into MHC binding and processing prediction tools. The former category contains tools that predict binding affinities of peptides to either MHC class I or class II molecules. Based on past benchmarks of MHC-I binding prediction methods, NetMHC-3.0 (called ANN on the IEDB website) and SMM methods are among the best performing individual methods (Peters et al., 2006; Lin et al., 2008a). For MHC-II, Consensus and NN-align have been shown to be best performing (Lin et al., 2008b; Wang et al., 2008). Consensus prediction methods, which combine outputs of different prediction approaches are also provided. Such methods have proven to be a successful strategy in many bioinformatics applications (Cuff et al., 1998; Ginalski et al., 2003; Arai et al., 2004).

The MHC class I processing predictions combine the binding predictions described above with predictions of proteasomal cleavage and TAP transport to identify peptides that are 'naturally processed'. Some studies have shown that the top predicted peptides are more likely to be recognized by T-cells than those selected by MHC binding predictions alone (Peters et al., 2003; Tenzer et al., 2005; Kessler and Melief, 2007). For example, in (Kessler and Melief, 2007), a number of integrated CD8+ epitope prediction methods available on the web were tested against a list of 64 nonamer peptides derived from a known tumor associated antigen, PRAME, that were previously studied. Out of 5 methods tested against this list, 4 methods identified 2 out of 3 known epitopes as among top 10 predicted peptides. The prediction method from the IEDB identified all 3.

However, no large-scale evaluation has been performed that clearly quantified the advantage of using the processing predictions, and determined how many epitopes are erroneously predicted not to be processed. Based on our experience to date, the advantage of using processing predictions over binding predictions -- if any -- seems to be relatively small. Also, any advantage has to be weighed against the increased complexity of interpreting the results of processing predictions, specifically when predicting epitopes in pathogens that have evolved to interfere with components of the processing machinery. In our own experimental applications we have therefore relied on MHC binding predictions alone, but we plan to revisit this following a systematical evaluation of the difference between the two approaches.

3 Application I: Determining known T-cell epitopes conserved in a novel strain of virus

RNA viruses such as Influenza rapidly mutate and can thereby escape detection from memory immune responses, allowing recurring infections of the same host by different strains from the same viral species. However, in every RNA virus there are some parts that mutate much less frequently, presumably because those parts are of high functional relevance, and most mutations result in a significant loss of fitness. The identification of epitopes that are conserved between different strains of a viral species is of particular interest. For both Dengue (Khan et al., 2008) and Influenza viruses (Heiny et al., 2007), known and predicted T-cell epitopes have been analysed for their conservation across multiple viral strains. A similar analysis was performed for the pandemic H1N1 strain in 2009 (Greenbaum et al., 2009), which was specifically probed for epitopes that were also present in recently circulating seasonal Influenza strains. A large fraction of T-cell epitopes was indeed found to be conserved in the pandemic strain, and these were experimentally confirmed to be targets of pre-existing T cell immunity.

The basic technical execution of identifying conserved T-cell epitopes from the IEDB in all of the above applications is similar. First, the IEDB is queried for epitopes from the viral species of interest using the 'source organism' field. For example, in FIGURE 1, 'Influenza A virus' is used as the 'source organism'. In addition, restrictions can be posed depending on the application, such as only T-cell epitopes could be of interest, or only those defined in humans. Further refinement of the query, such as restricting epitopes to those mapped following infection with virus (in contrast to immunization with individual antigens), requires using the advanced search option. Once the set of epitopes has been retrieved from the IEDB, the sequences can be exported as an Excel spreadsheet, and analyzed for conservation offline. This can either be done manually, or with custom computer code, or using the IEDB epitope conservancy tool (Bui et al., 2007). Using the epitope conservancy tool, the retrieved epitopes can be checked for conservation in different antigen sequences. For instance, by entering the protein sequences of the 2009 pandemic Influenza strain (see FIGURE 3), and selecting a sequence identity threshold of 100%, all previously known Influenza epitopes that are 100% conserved in the pandemic strain can be identified.

For individual antigens, a simple alternative to search for all known epitopes that are conserved in it is to enter the antigen protein sequence in the 'Linear Peptide' field (see FIGURE 1), and selecting the 'Blast 90%' option rather than 'Exact Matches' in the pull down menu to retrieve partially matching epitopes meeting the similarity threshold. This query can again be combined with further restrictions, such as limiting the results to epitopes recognized by T-cells, defined in humans and derived from Influenza A virus.

4 Application II: Identifying novel T-cell epitope targets to facilitate vaccine design

The purpose of identifying T-cell epitopes for vaccine design can be multifold. Sets of epitopes can be used directly as a vaccine (Sette and Fikes, 2003; Kessler and Melief, 2007; Purcell et al., 2007), or epitope mapping can be used to identify antigens that should be included in a subunit vaccine (Scarselli et al., 2005; Zvi et al., 2008), or epitopes can be used as a correlate of protective efficacy to evaluate different vaccine constructs. The task at hand in each of these cases is to identify peptides in a given antigen that elicit a T-cell response. In contrast to the applications above, we here consider the identification of novel epitopes in antigens that have not been studied exhaustively. For vaccine design applications that involve well-studied pathogens, the query strategies described in the previous section should be applied. To identify novel epitopes, a multi-step approach has been shown to be successful, in which prediction tools are used to identify epitope candidates that are then tested in validation experiments. The goal of predictions is therefore to identify a small enough set of peptide candidates that make experimental validation feasible.

When making peptide:MHC binding predictions, there are a number of important things to consider to achieve optimal results. First, the length of peptides that are considered to be potential ligands are different for MHC class I and II molecules. Most MHC class I restricted T-cell epitopes are 9 residues long, but depending on the MHC allele, epitopes of lengths 8, 10 and 11 are also found. That means, for a given antigen, all peptides of these lengths are potential candidate epitopes. In contrast, MHC class II restricted T-cell epitopes are of more variable lengths between 13 and 25 residues, as they are bound in an open ended groove and ends of a bound peptide can be outside of the core binding region. The core region of six to nine epitope residues makes direct contact with the binding pocket of the MHC class II molecule. Peptides containing the same core region with flanking residues of different length can activate the same T-cells, arguably making them variants of the same epitope. However, the core binding region by itself is a poor epitope, as peptides of length 9 or shorter do not bind MHC class II well. A peptide overhang of 2-3 residues at each end is

required to stabilize the peptides, largely independent of what residues are involved. This means 13-mers or 15-mers are more appropriate to test as MHC class II epitope candidates. In our experimental applications, we have historically used 15-mer peptides.

Second, MHC alleles for which peptide binding predictions will be made need to be carefully chosen. If the goal of the study is to identify T-cell epitopes for use in an animal model, the selected MHC alleles should match those of the species and strain of animals used. If the goal is to design a vaccine for a human population, the frequency of MHC alleles in the population of interest should be considered to maximize the vaccine's population coverage. As every human expresses multiple MHC class I and II molecules, it is not required that a vaccine covers all of them, but at least one MHC allele in the majority of individuals should be included. There are different approaches for achieving this goal. For instance, it has been noted that binding specificities of MHC-I molecules overlap and can be grouped into 9 supertypes (Sidney et al., 2008). By selecting alleles representing several or all of these supertypes, one can achieve a wide coverage. Alternatively, a collection of MHC alleles can be chosen and the population coverage of a set of epitopes reactive for each of these alleles can be projected using the Population coverage tool (Bui et al., 2006).

As an example application, we will consider identification of T-cell epitopes in Sabia virus restricted by two prototype supertype alleles, HLA A*0201 and HLA A *1101. This is a simplified version of the strategy in (Kotturi et al., 2009) in which a crossreactive vaccine for multiple Arenavirus species was designed. First, one needs to obtain protein sequences from Sabia virus, which can be done by searching the NCBI database or a specialist resource such as the NIAID Bioinformatics Resource Centers (Greene et al., 2007) for full genomic sequences of Sabia. The virus has two genomic segments encoding four proteins with a total of 3362 residues (in this example, the proteins with genbank identifiers [gi|52627080](#), [gi|52627081](#), [gi|52627077](#) and [gi|52627078](#) were used). These sequences are then pasted in FASTA format into the MHC class I binding prediction tool. Selecting 'HLA A*0201' as the allele to make binding predictions for, and 'all lengths' to make predictions for and 'submit' results in 6,656 predictions of affinities for HLA A*0201 for all 9 and 10-mer peptides contained in the proteins. Multiple prediction methods are available, but by default the method considered best is used, which currently is the 'ann' method (NetMHC, (Lundegaard et al., 2008)). The peptides can be sorted by IC50 value with low values corresponding to high affinity binders. The same can now be done for HLA A*1101, resulting in a second list of peptide candidates sorted by their predicted affinity.

Once the sorted lists of peptide candidates are available, the most promising candidates for validation experiments can be chosen using a number of strategies. One strategy is to choose the top 0.5% - 3% scoring peptides for each allele, thereby equally representing all alleles. Another strategy is to use the same IC50 cutoff for all alleles, which will result in different number of peptides for alleles, because of different sizes of binding peptide repertoires. What strategy is preferred depends on the application: If it is more important to identify one or more epitopes per allele, one should represent the alleles equally, and use a percentile cutoff. If it is more important to maximize the chances of getting a response per peptide, independent of the restricting allele, we believe it is better to choose an absolute IC50 value cutoff. Still another is based on the use of a promiscuity score, which is defined as the number of alleles that a peptide binds at a given binding affinity score threshold. Using this strategy, peptides with higher promiscuity scores would be selected. We have been primarily using the first strategy in order to have unbiased representation of each allele in the candidate set. We are planning to perform a meta analysis of the results from these studies to determine if different strategies of peptide selection would have resulted in higher success rates.

The top 67 peptides (=1%) of all predicted binders should include more than 80% of all epitopes restricted by the allele of interest in a given virus (Moutaftsi et al., 2006; Kotturi et al., 2007). This number of peptides is also easily testable in validation experiments. Validation experiments should typically be carried out in a staggered fashion, where high throughput experiments are carried out first to further narrow the set of candidates to be tested in more resource intensive experiments. In (Kotturi et al., 2009), predicted binding peptides from Arenaviruses were first tested for recognition by T-cells from HLA transgenic mice that were infected with recombinant vaccinia virus (rVACV) expressing Arenavirus antigens. The use of HLA transgenic mice rather than samples from previously infected humans was necessary as the latter are very hard to obtain. Similarly, the use of rVACV was necessary as using live Arenavirus infections poses substantial biosafety concerns. Peptides that were targeted by strong responses were selected for further validation, carried out by immunizing HLA transgenic mice with the peptide, and then challenging them with a lethal dose of recombinant rVACV. Ultimately, a set of epitopes was assembled that resulted in protection against different strains of Arenavirus, demonstrating cross-protectiveness of the vaccine. The HLA-A*0201 restricted Sabia virus peptides included in the vaccine were #2 and #37 highest predicted binders (GLLEWIFRA and LLPDALLFTL) in the candidate list generated as described above. This illustrates that a significant fraction of peptide candidates will not be able to induce effective protection, even though they had similar or better predicted binding affinities to the ones that did. This always has to be expected, as a multitude of additional factors influence the ability of an epitope to protect, such as availability of a T cell repertoire, or kinetics of antigen expression.

5 CONCLUSION

The IEDB (<http://www.iedb.org>) contains useful resources for those studying the immune system and developing immune-based vaccines, therapeutics and diagnostics. This review has shown examples demonstrating how T-cell immune epitope data and epitope prediction tools provided by the IEDB can be effectively used. We expect that, as epitope based applications are increasingly used in a high throughput manner, new tools will become available that better predict what epitopes are most appropriate for different applications. Also, we are continuously working on improving the usefulness of the IEDB resource, in particular in further automating the workflow of novel epitope identification. We are grateful for the feedback we have received from the IEDB user community, and want to further encourage constructive criticism, as ultimately the success of the IEDB is measured by the value it provides to its user base.

Acknowledgments

This work was supported by the National Institutes of Health's (National Institute of Allergy and Infectious Disease) Contract HHSN26620040006C.

REFERENCES

- Altman JD, Moss PAH, Goulder PJR, Barouch DH, McHeyzer-Williams MG, Bell JI, McMichael AJ, Davis MM. Phenotypic Analysis of Antigen-Specific T Lymphocytes. *Science*. 1996; 274:94–96. [PubMed: 8810254]
- Arai M, Mitsuke H, Ikeda M, Xia J-X, Kikuchi T, Satake M, Shimizu T. ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability. *Nucl. Acids Res*. 2004; 32:W390–393. [PubMed: 15215417]
- Banchereau J, Palucka AK. Dendritic cells as therapeutic vaccines against cancer. *Nat Rev Immunol*. 2005; 5:296–306. [PubMed: 15803149]

- Beaver J, Bourne P, Ponomarenko J. EpitopeViewer: a Java application for the visualization and analysis of immune epitopes in the Immune Epitope Database and Analysis Resource (IEDB). *Immunome Research*. 2007; 3:3. [PubMed: 17313688]
- Bousquet J, Lockey R, Malling H-J. Allergen immunotherapy: Therapeutic vaccines for allergic diseases A WHO position paper. *Journal of Allergy and Clinical Immunology*. 1998; 102:558–562. [PubMed: 9802362]
- Brock I, Weldingh K, Leyten EMS, Arend SM, Ravn P, Andersen P. Specific T-Cell Epitopes for Immunoassay-Based Diagnosis of Mycobacterium tuberculosis Infection. *J. Clin. Microbiol*. 2004; 42:2379–2387. [PubMed: 15184408]
- Bryson CJ, Jones TD, Baker MP. Prediction of Immunogenicity of Therapeutic Proteins: Validity of Computational Tools. *BioDrugs*. 2010; 24
- Bui H-H, Sidney J, Dinh K, Southwood S, Newman M, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2006; 7:153. [PubMed: 16545123]
- Bui H-H, 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; 8:361. [PubMed: 17897458]
- Cuff J, Clamp M, Siddiqui A, Finlay M, Barton G. JPred: a consensus secondary structure prediction server. *Bioinformatics*. 1998; 14:892–893. [PubMed: 9927721]
- Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics*. 2003; 19:1015–1018. [PubMed: 12761065]
- 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. *Proceedings of the National Academy of Sciences*. 2009; 106:20365–20370.
- Greene JM, Collins F, Lefkowitz EJ, Roos D, Scheuermann RH, Sobral B, Stevens R, White O, Di Francesco V. National Institute of Allergy and Infectious Diseases Bioinformatics Resource Centers: New Assets for Pathogen Informatics. *Infect. Immun*. 2007; 75:3212–3219. [PubMed: 17420237]
- Heiny AT, Miotto O, Srinivasan KN, Khan AM, Zhang GL, Brusic V, Tan TW, August JT. Evolutionarily Conserved Protein Sequences of Influenza A Viruses, Avian and Human, as Vaccine Targets. *Plos One*. 2007; 2:e1190. [PubMed: 18030326]
- Jones T, Crompton L, Carr F, Baker M. Deimmunization of monoclonal antibodies. *Methods Mol Biol*. 2009; 525:405–423. [PubMed: 19252848]
- Kaech SM, Wherry EJ, Ahmed R. Effector and memory T-cell differentiation: implications for vaccine development. *Nat Rev Immunol*. 2002; 2:251–262. [PubMed: 12001996]
- Kessler JH, Melief CJM. Identification of T-cell epitopes for cancer immunotherapy. *Leukemia*. 2007; 21:1859–1874. [PubMed: 17611570]
- Khan AM, Miotto O, Nascimento EJM, Srinivasan KN, Heiny AT, Zhang GL, Marques ET, Tan TW, Brusic V, Salmon J, August JT. Conservation and Variability of Dengue Virus Proteins: Implications for Vaccine Design. *Plos Neglected Tropical Diseases*. 2008; 2:e272. [PubMed: 18698358]
- Korber, BTM.; Brander, C.; Haynes, BF.; Koup, R.; Moore, JP.; Walker, BD.; Watkins, DI. Los Alamos National Laboratory, Theoretical Biology and Biophysics. Los Alamos; New Mexico: HIV Molecular Immunology 2006/2007.. LA-UR 07-4752
- Kotturi MF, Botten J, Sidney J, Bui H-H, Giancola L, Maybeno M, Babin J, Oseroff C, Pasquetto V, Greenbaum JA, Peters B, Ting J, Do D, Vang L, Alexander J, Grey H, Buchmeier MJ, Sette A. A Multivalent and Cross-Protective Vaccine Strategy against Arenaviruses Associated with Human Disease. *PLoS Pathog*. 2009; 5:e1000695. [PubMed: 20019801]
- Kotturi MF, Peters B, Buendia-Laysa F Jr. Sidney J, Oseroff C, Botten J, Grey H, Buchmeier MJ, Sette A. The CD8+ T-Cell Response to Lymphocytic Choriomeningitis Virus Involves the L Antigen: Uncovering New Tricks for an Old Virus. *J. Virol*. 2007; 81:4928–4940. [PubMed: 17329346]

- Lin H, Ray S, Tongchusak S, Reinherz E, Brusic V. Evaluation of MHC class I peptide binding prediction servers: Applications for vaccine research. *BMC Immunology*. 2008a; 9:8. [PubMed: 18366636]
- Lin H, Zhang G, Tongchusak S, Reinherz E, Brusic V. Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. *BMC Bioinformatics*. 2008b; 9:S22. [PubMed: 19091022]
- Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucl. Acids Res*. 2008; 36:W509–512. [PubMed: 18463140]
- Moutaftsi M, Peters B, Pasquetto V, Tschärke DC, Sidney J, Bui H-H, Grey H, Sette A. A consensus epitope prediction approach identifies the breadth of murine TCD8⁺-cell responses to vaccinia virus. *Nat Biotech*. 2006; 24:817–819.
- Pai M, Zwerling A, Menzies D. Systematic Review: T-Cell based Assays for the Diagnosis of Latent Tuberculosis Infection: An Update. *Annals of Internal Medicine*. 2008; 149:177–184. [PubMed: 18593687]
- Peters B, Bui H-H, Frankild S, Nielsen 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; 2:e65. [PubMed: 16789818]
- Peters B, Bulik S, Tampe R, van Endert PM, Holzhutter H-G. Identifying MHC Class I Epitopes by Predicting the TAP Transport Efficiency of Epitope Precursors. *J Immunol*. 2003; 171:1741–1749. [PubMed: 12902473]
- Peters B, Sette A. Integrating epitope data into the emerging web of biomedical knowledge resources. *Nat Rev Immunol*. 2007; 7:485–490. [PubMed: 17479127]
- Peters B, Sidney J, Bourne P, Bui H-H, 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; 3:e91. [PubMed: 15760272]
- Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov*. 2007; 6:404–414. [PubMed: 17473845]
- Sathiamurthy M, Peters B, Bui H-H, Sidney J, Mokili J, Wilson S, Fleri W, McGuinness D, Bourne P, Sette A. An ontology for immune epitopes: application to the design of a broad scope database of immune reactivities. *Immunome Research*. 2005; 1:2. [PubMed: 16305755]
- Scarselli M, Giuliani MM, Adu-Bobie J, Pizza M, Rappuoli R. The impact of genomics on vaccine design. *Trends in Biotechnology*. 2005; 23:84–91. [PubMed: 15661345]
- Scott DW, De Groot AS. Can we prevent immunogenicity of human protein drugs?. 2010;i72–i76.
- Sette A, Fikes J. Epitope-based vaccines: an update on epitope identification, vaccine design and delivery. *Current Opinion in Immunology*. 2003; 15:461–470. [PubMed: 12900280]
- Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. *BMC Immunology*. 2008; 9:1. [PubMed: 18211710]
- Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel P-M, Rammensee H-G, Schild H, Holzhutter H-G. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. 2005:1025–1037.
- Vita R, Peters B, Sette A. The curation guidelines of the immune epitope database and analysis resource. *Cytometry*. 2008; 73A:1066–1070. [PubMed: 18688821]
- Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B. The Immune Epitope Database 2.0. *Nucl. Acids Res*. 2010; 38:D854–862. [PubMed: 19906713]
- Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. *PLoS Comput Biol*. 2008; 4:e1000048. [PubMed: 18389056]
- Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, Bui H-H, Buus S, Frankild S, Greenbaum J, Lund O, Lundegaard C, Nielsen M, Ponomarenko J, Sette A, Zhu Z, Peters B. Immune epitope database analysis resource (IEDB-AR). *Nucl. Acids Res*. 2008; 36:W513–518. [PubMed: 18515843]

Zvi A, Ariel N, Fulkerson J, Sadoff J, Shafferman A. Whole genome identification of Mycobacterium tuberculosis vaccine candidates by comprehensive data mining and bioinformatic analyses. BMC Medical Genomics. 2008; 1:18. [PubMed: 18505592]

Search

Epitope Structure

☐ Any

☒ Linear Peptide:

Exact Matches ▾

☐ Discontinuous Peptide

☐ Non-Peptidic:

Molecule Finder

Epitope Source

Source Organism: Influenza A virus;

Organism Finder

Source Antigen:

Molecule Finder

Immune Recognition Context

☐ B Cell Response

☒ T Cell Response

☐ MHC Binding

☐ MHC Ligand Elution

Host Organism: Homo sapiens;

Organism Finder

MHC Restriction:

Allele Finder

MHC Class: II ▾

Search

Clear

FIGURE 1.

Main IEDB query interface for retrieving epitopes. Using the interface, one can specify epitope structure type (Epitope Structure), organism from which it came from (Epitope Source), and contexts in which the epitope was determined (Immune Recognition Context). An example query for retrieving all T-cell epitopes derived from Influenza A virus and restricted to human host and MHC class II is shown.

A

Epitopes	Positive *	Negative **	All
Peptidic	276	81	357
Non-Peptidic	0	0	0

Assays	Positive	Negative	All
T Cell Response	524	103	627
B Cell Response	0	0	0
MHC Ligand Elution	0	0	0
MHC Binding	0	0	0

Summary	
Epitope Source Organism	23
Host Organism	4
Restricting MHC Allele	52
References	53

B

T Cell ID ↑	Reference	Epitope	Host	Immunization	Assay Antigen	Antigen Epitope Relation	MHC Restriction	Assay Description
16823	T Linnemann, J Virol 2000	FVFTLTVPSE Matrix protein 1 (62-72) Influenza A virus (A/Puerto Rico/8/1934(H1N1))	Homo sapiens	Infectious disease via exposure to Influenza A virus (Taxonomic Parent) followed by restimulation in vitro	FVFTLTVPSE Matrix protein 1 (62-72) Influenza A virus (A/Puerto Rico/8/1934(H1N1))	Epitope	HLA-DR4	51 Chromium Release Killing Positive
17773	Ming Ye, Hum Immunol 2004	PKYVKQNTLKLAT Hemagglutinin precursor (306-318) Influenza A virus	Homo sapiens	Exposure to endemic/ubiquitous agent Influenza A virus (Source Organism) without evidence for disease followed by restimulation in vitro	PKYVKQNTLKLAT Hemagglutinin precursor (306-318) Influenza A virus	Epitope	HLA-DRB1*0401	MHC Tetramer/Multimer Staining TCR Binding Positive
17774	Ming Ye, Hum Immunol 2004	PKYVKQNTLKLAT Hemagglutinin precursor (306-318) Influenza A virus	Homo sapiens	Exposure to endemic/ubiquitous agent Influenza A virus (Source Organism) without evidence for disease followed by restimulation in vitro	PKYVKQNTLKLAT Hemagglutinin precursor (306-318) Influenza A virus	Epitope	HLA-DRB1*0401	MHC Tetramer/Multimer Staining TCR Binding Positive

FIGURE 2.

Epitope query results. Using the query shown in Figure 1, the two panels show the query result at different levels of detail. Panel A provides a summary, which is presented immediately after the query is made. In the panel, the first table indicates how many unique epitope structures are 'Positive' or 'Negative'. Here 'Positive' is used to mean at least one positive assay matched the search criteria. The second table indicates number of assay results for the epitopes retrieved. Because the same epitope may have been studied in multiple assays, the numbers will be larger than those in the first table. Clicking the hyperlink in the second table for 'Positive' T-cell response results in panel B, which shows a sample of three entries from the detailed list of all positive T-cell assay results associated with the query. In panel B, in addition to listing epitope structures, details of assays such as references that described the experiment, immunization, MHC restriction, source antigens, and assay description, among others, are provided.

Step 1. Epitope Sequence(s)	
Enter epitope sequence(s) in PLAIN or FASTA format: ?	<div></div>
Or upload epitope sequence(s) from a file: ?	<div>Choose File No file chosen</div> <div>Click here to upload</div>
Step 2. Protein Sequence(s)	
Enter protein sequence(s) in PLAIN or FASTA format: ?	<div></div>
Or upload protein sequence(s) from a file: ?	<div>Choose File No file chosen</div> <div>Click here to upload</div>
Click here to browse for sequences in NCBI ?	
Step 3. Calculation option(s)	
Analysis type: ?	☒ Epitope linear sequence conservancy ☐ Epitope discontinuous sequence conservancy
Sequence identity threshold: ?	<div>≥ 100%</div>
Remove duplicated protein sequences? ?	☐ Yes
Step 4. Submission	
<div>Submit Reset</div>	

FIGURE 3.

Epitope conservancy tool. The tool allows the user to determine to what extent user-provided epitopes are conserved in a list of protein sequences. The two text input boxes allow the user to enter epitopes and protein sequences of interest, respectively. After specifying the type of epitope structure (i.e. linear vs. discontinuous) and sequence identity threshold and submitting, results are presented that indicate how many proteins contain each epitope at the specified threshold. Additional analysis results such as alignments of epitopes and proteins are also provided.

TABLE 1

Main curation categories of peptidic entries in the IEDB

Main Category	Total number journal articles	% Processed
Infectious Diseases	8616	99
Allergens	741	97.7
Autoimmune	3924	84.1
Transplants/ Alloantigens	631	20.4

For each category, a broad PubMed query was carried out to retrieve all journal articles of potential relevance. These article abstracts were then scanned using a combination of automated and manual text classification to filter out false positives, resulting in the numbers shown in the second column. Any epitopes from these articles were then manually curated into the IEDB. The percentage completion of the manual curation process in each category to date is shown in the last column.

TABLE 2

Categories of T-cell epitope prediction and analysis tools available at the IEDB

Tools Type	Category	Description
Prediction	MHC-I	Predict peptides that bind to MHC class I molecules
	MHC-II	Predict peptides that bind to MHC class II molecules
	Processing (MHC-I)	Combine proteasome and TAP binding predictions with MHC-I binding ones
Analysis	Homology Mapping	Map epitopes onto 3D protein structure of their source antigen or its homologue
	Epitope Cluster Analysis	Group epitopes based on their sequence similarities
	Epitope Conservancy Analysis	Calculate degree of conservation of epitopes in a set of provided proteins
	Population Coverage	Estimate population coverage of a set of epitopes based on HLA frequencies