Determining Potential Epitopes for Vaccine Candidacy Using IEDB

Why are epitopes important and how can IEDB help us search for them?

Epitope binding of MHC is the most selective step during immune response. It determines whether peptides cleaved from a foreign protein by the proteasome will be degraded in the ER after being transported there by TAP, or whether they will complex with MHC molecules successfully. If binding affinity is high enough, the MHC-peptide complex will be transported through the Golgi to the cell surface, where MHC will present this epitope to T-cells. The T-cells will bind it and search for antigens containing the epitope in order to destroy them.

Identifying what potential epitopes will bind with the greatest affinity to the MHC allows us to find which ones will most likely mobilize the immune system against a pathogen. It helps for vaccine development, which can include peptides that has been identified as effective epitopes, in order to bind MHC and enable T cells to identify the antigen they need to bind to and destroy to protect against the disease in question.

IEDB is an abbreviation for the Immune Epitope Database and Analysis Resource. It is a website that enables the user to identify epitopes - known, as well as predicted – that will most likely be presented to antibodies in a host by determining which of these peptides bind specific MHC Class I and II alleles most effectively. It also aggregates epitope data for specific host-disease interactions that have been discovered in research. The functionality of IEDB can be represented in two halves: The first is a robust search engine accessing known epitope data. IEDB allows the user to search a rich database of captured experimental data from a library of curated journal articles. The second is a suite of computational methods that predict epitopes and their immunogenicity. These methods enable selection of epitopes for integration in vaccines by predicting antigen processing (proteosomal cleavage and TAP binding), binding between peptides and MHC Class I and II molecules, as well as T- cell immunogenicity scores. With these two major components, IEDB gives researchers a tool to analyze immune interactions in many contexts and leverage this knowledge to fight disease.

A short list of applications of IEDB:

1. To search all known epitopes for a host and disease, with links to research papers they have been identified in.
2. To determine the dynamics of an antigen-epitope relationship, with links to research papers
    - What is the level of immunogenicity of this epitope? Was it protective against the antigen or not?
3. To search using an epitope sequence to figure out what protein it originates from, its position on that protein, the host it is associated with, and the properties of immunization of that host in experimental research.
4. It enables the user to find all known epitopes of an organism and its **antigens** that you indicate in the search, for the specific type of immune response you are interested in (T-cell response or B-cell response)

- You indicate this in your search by filtering by **assay** (T-cell assay, B-cell assay, etc.)
- You can also specify in the assay filter parameters that you are searching for epitopes involved in "**treatment**" of the disease you've indicated, which will search applicable research papers using this context
- Results will have different hosts that you can sort by, or you can specify a single host to filter by before you search

5. To visualize epitopes of an antigen (protein) by mapping them to a reference protein (done using Immunome Browser), determining the response frequency involving the antigen in question based on number of positive and negative assays related to total number of subjects tested (from the literature)
6. To determine potential MHC-binding ligands in a given protein sequence that bind to a specific MHC Class I or II allele, using computational methods to rank these epitopes according to their binding affinity with MHC, as well as efficiency of processing by the proteasome and TAP transport that precede binding.

What can IEDB do that other epitope databases and predictive tools can't?

IEDB holds some distinct advantages over older tools that came before it, which in comparison, are less comprehensive, offer less accuracy from their predictive tools, and do not boast a powerful search engine. Unlike its forbears, SYFPEITHI and BIMAS, IEDB offers a large suite of the latest, high-tech prediction methods. These methods employ machine learning methods that can evaluate how sequence context affects binding affinity using artificial neural networks and support vector machines. One example is NetMHCpan, which predicts novel epitopes, not only filtering which results have the greatest chances of eliciting an immune response but doing so with lower rates of error. To add to this advantage, IEDB utilizes its enormous catalog of assay-derived peptide binding information in order to train these binding prediction tools. One example of how these methods outpace the older predictive models used in BIMAS and SYFPEITHI is the ability of NetMHC to use 9-mer binding data (the most common peptide length in MHC Class I binding) in order to infer binding affinities for ligands of 8,10, and 11 residues. NetMHC and NetMHCpan have both been proven to be more accurate than BIMAS and SYFPEITHI methods, with the former group resulting in higher Pearson correlation coefficients (between predicted and observed binding) than the latter group (Lundegaard et al., 2010, p. 313).

IEDB does not only focus on MHC binding, however. It also generates processing scores that involve proteasomal cleavage and TAP transport, in order to give researchers additional filters to narrow down long lists of potential epitopes they are considering (many times MHC-binding affinity calculations are enough, but other times the results may be too broad and require further levels of categorization and analysis using processing calculation methods).

Using the Population Coverage Calculation tool a user can input these epitopes and narrow down which have the greatest binding affinities for MHC class I or II alleles common in a specific population (say, the United States), in order to target a vaccine to these individuals. Alternatively, we can figure out which of these epitopes have the greatest applicability in terms

of covering multiple populations, for the purpose of developing a vaccine that can be administered across regions or even worldwide.

The Immunome Browser is another unique tool from IEDB that enables the visualization of immunogenic residues on an antigen or antigen-containing protein. It maps each known epitope of an antigen to specific positions on the reference protein (either the antigen itself or the larger protein it is a component of) categorizing each by response frequency (amount of positive responses divided by the total number of responses) according to research papers that involve these epitopes. It is a useful way to summarize the overall immunogenic character of an antigen.

IEDB isn't the only database that offers immune-related information, having to share a space with other tools such as AntiJen, EPIMHC, MHCBN, and Vaxijen that offer similar search functionality. AntiJen hosts data on T-cell and B-cell epitopes, while EPIMHC has data on T-cell and MHC epitopes. Meanwhile, MHCBN presents data on TAP binding affinities in addition to T-cell and MHC epitopes. Each of these tools lacks certain search categories that IEDB covers, however. None of the three present searchable data for each of these categories in one location, except for IEDB (Flower & Perry, 2013, pgs. 51, 122, 129).

The IEDB search engine does miss a few types of immune data though. Antijen offers kinetic and thermodynamic data as well as quantitative data on molecular interactions between epitopes and proteins as well as proteins and proteins in the context of the immune system, something that is missing from IEDB. Vaxijen, on the other hand, allows users to identify potential protective antigens using its database, which is an additional search functionality not found on IEDB (Flower & Perry, 2013, pgs. 51, 56, 122, 129).

Researchers have also identified a few potential drawbacks of using IEDB epitope prediction. The prediction tools that IEDB offers use algorithms that consider a very large number of immunoassay methods, sometimes resulting in data noise that makes results less reliable (Lundegard et al., 2010, pp. 311). Additionally, though MHC molecules are classified into supertypes of alleles that overlap peptidically, these classifications sometimes result in inaccuracy during epitope prediction because it is still not guaranteed that these alleles have the same binding motifs. For this reason, only about 100 of 2000 total MHC alleles can be used to make allele-specific predictions. IEDB attempts to solve this problem using NetMHCpan, which uses allele-specific binding data in order to evaluate peptide binding with MHC alleles that have little or no previously determined binding data. These calculations are performed using the sequence similarity distance between alleles with little or no data and alleles that do have enough data (Lundegaard et al., 2010, p.313).

An example dataset for testing the IEDB search and binding affinity prediction tools:

Subunit vaccines containing immunogenic antigens have been proposed for use

in tandem with the M. bovis BCG vaccine, in order to boost its efficacy (Stylianou et al., 2018, p. 3). In the study "Identification and Evaluation of Novel Protective Antigens for the Development of a Candidate Tuberculosis Subunit Vaccine", researchers discovered that the antigen PPE15 elicited a robust immune response and suggested that it could be incorporated in a booster vaccine to ensure a higher rate of immunization. This study was done in mice (Stylianou et al., 2018, p. 11). PPE15 has a UniProt ID of P9WI31 and a Gene ID of Rv1039c.

The dataset includes 7 peptides (found using the IEDB search engine) associated with PPE15, and binding affinity will be explored with seven DRB1 and DRB3/4/5 alleles (MHC Class II) whose peptides have been cited by IEDB as being the most immunodominant worldwide. The peptides studied were derived from 15 allergen and bacterial antigens and bind with the 26 most common HLA Class II alleles in the world (Paul et al., 2015, p. 28). To narrow down these alleles to 7, researchers calculated the lowest average percent of peptides (of the total peptides studied) determined to be responsible for half of total antigen-specific responses recorded. The peptides that make up this group were determined to be the most immunodominant peptides (Paul et al., 2015, p. 31). By using the MHC Class II alleles that bind these peptides, users can predict the epitopes most widely applicable to the world population, using the lowest number of alleles to achieve this coverage (the seven alleles listed above).

I want to figure out what human epitopes of this antigen are predicted by IEDB, specifically, which peptides meet the 20% threshold for binding affinity (top 20% of all scores). This 20 percent threshold was determined by using IEDB prediction tools to score binding between peptides generating half of antigen-specific responses in the experiment and the seven MHC Class II alleles that were determined above. The median of consensus percentile ranks from each peptide (based on binding affinity score) was taken, which was 20.0 (Paul et al., 2015, p.32).

In another application I will be predicting epitopes using the MHC Class II binding prediction tool as well as predicting hypothetical MHC ligands (using a tool that calculates cleavage probability) from PPE15, using its amino acid sequence:

MDFGALPPEINSARMYAGAGAGPMMAAGAAWNGLAAELGTTAASYESVITRLTTESWMGP
ASMAMVAAAQPYLAWLTYTAEAAAHAGSQAMASAAAYEAAYAMTVPPEVVAANRALLAAL
VATNVLGINTPAIMATEALYAEMWAQDALAMYGYAAASGAAGMLQPLSPPSQTTNPGGLA
AQSAAVGSAAATAAVNQVSVADLISSLPNAVSGLASPVTSVLDSTGLSGIIADIDALLAT
PFVANIINSAVNTAAWYVNAAIPTAIFLANALNSGAPVAIAEGAIEAAEGAASAAAAGLA
DSVTPAGLGASLGEATLVGRLSVPAAWSTAAPATTAGATALEGSGWTVAAEEAGPVTGMM
PGMASAAKGTGAYAGPRYGFKPTVMPKQVVV

For this test I will be using HLA- DRB1*03 and HLA-DRB1*07 as my MHC Class II alleles, with a peptide-length range of 12-18 amino acids (the IEDB default). In Brazil, both Caucasian and Afro-Brazilian populations exhibit high frequencies of haplotypes containing DRB1*03 and DRB1*07 alleles (Bardi, Jarduli, Jorge, Camargo, Carneiro, Gelinksi, Silva, & Lavado, 2011, p. 5).

MANUAL

1. Searching IEDB for known MHC Class II epitopes of an antigen of interest

Step 1: User should identify a pathogen that they intend to target for vaccination. Once this pathogen is chosen, scientific research on known antigens of this pathogen should be reviewed. Choose an antigen of interest that is hypothesized to be immunogenic. Note down the UniProt ID for this antigen, as it will be used later in this tutorial.

Step 2: Navigate to the IEDB home page in order to begin a search. In order to search broadly, select “Linear Epitopes” under *Epitope* and “Humans” under *Host*, then click “Search”.

Step 3: All known linear epitopes in the database that have positive immunoassay results in humans are displayed here. In a column on the left side of the web page there are additional filters for results. Find the *Antigen* filter and next to “Antigen Name” click the “Finder”. A popup appears where users can search for their antigen. User should enter the UniProt ID of the antigen into the “Molecule ID” field, then hit “Search”.

Step 4: Search results for the antigen will appear at the bottom of the pop-up window. Users can find their antigen by name, then click the green “+” sign next to it, then hit “Apply” to add this antigen to their search parameters.

Step 5: Results will be narrowed down to known epitopes of the antigen of interest. If there are no results, user should check to see what filters are currently in use. These can be found above the results under “Current Filters”. Usually, unchecking “Positive Assays Only” will allow the user to view peptides of the antigen that are not known by IEDB to have immunogenic qualities. However, these epitopes may still have relevancy. While IEDB has access to over 2 million experiments and 20,000 journal articles, it is possible that research papers not yet integrated into the database indicate the existence of immunogenic epitopes from this antigen (Martini, Nielsen, Peters, Sette, 2019, p. 4). Once the unwanted filter (or filters) is unchecked, hit “Search”.

Step 6: If there are still no results, then this means that IEDB has not yet curated manuscripts that contain epitope information for the antigen of interest. If there are epitope results, more information about the results can be determined through the search page. First, note the length of the epitopes if you do not know what family of MHC molecules they bind with. MHC Class I molecules typically bind with peptides with a length of 8 to 10 residues, while it is common for MHC Class II molecules to bind with peptides of 12 to 24 residues (Cruz-Tapias, Castiblanco ,& Anaya, 2013, pp. 175-176).

Step 7: Click on the assay tab in order to see what assays were used in evaluating each listed epitope. The “Assay Description” column indicates what method was used to measure immune cell response in the experiment and whether the result was “Positive” or “Negative”. On the opposite side, under the “ID” column, click the links to examine reference information for this epitope. Here there is more detail about the title, authors, and year, as well as the abstract of the paper and various epitopes and immunoassays that may have relevance to this research. This gives the user a quick way to check a particular epitope for relevance, or to dig deeper into research involving this epitope or the antigen it originates from.

Step 8: Returning to the results page, look for the “Immunization” column. This indicates the method by which antigenic material was introduced to cells in the host organism, providing additional context for the user to interpret the search results.

Step 9: Clicking on the “Reference” tab, the user can view a page that includes similar information to the previously viewed “Reference Information” page, but additionally includes PMIDs (PubMed IDs) of the manuscripts from which IEDB captured each epitope in the search results. By clicking a PMID, the user can quickly and easily access these journal articles through the PubMed website.

Note: there is not always a corresponding PubMed ID for each result, in which case the user must manually search online in order to find the manuscript of interest).

2. Verifying MHC Class II epitope immunogenicity to filter candidates for use in vaccines

Step 1: User should copy each epitope sequence from the search results to their clipboard. After the sequences are saved to the clipboard, the user should return to the IEDB mainpage, look for *Epitope Analysis*, and click on either “MHC II Binding” or “MHC I Binding” under “T-Cell Epitope Prediction”, depending on what type of epitopes the user is analyzing.

Note I: To keep this manual succinct, the next steps will detail the use of the MHC II Binding prediction tool only, as the MHC I binding prediction tool is highly similar in features.

Note II: In other cases, the user may want to use the full protein sequence of an antigen rather than epitope sequences found using the IEDB search tool, in order to calculate binding predictions.

Step 2: User should paste the copied epitope sequences into the “Enter protein sequences” field. The user also has the option of copying and pasting FASTA file content as well as uploading an entire FASTA file into this field.
When choosing a prediction method under “Prediction Method”, IEDB recommends that users select “IEDB recommended 2.22”. This allows algorithms to determine what prediction method or methods (NetMHCIIpan, SMMalign, NNalign, Sturniolo, Combinatorial Library, or a consensus of various methods) will most accurately predict the binding of the user specified epitopes and MHC Class II alleles.

Under “Select species/locus” the user can choose between various HLA class II supertypes they are interested in examining (this includes HLA-DP, HLA-DQ, and HLA-DR). Once the supertype is chosen, the user can narrow down alleles of interest by adding alleles or removing them with the delete button. Alternatively, there are three other options for specifying HLA alleles. The first is a “7-Allele HLA reference set”, which can be used to screen for epitopes most widely applicable to the world population. The second is a full HLA reference set (from which the user can narrow alleles by interest). The last option is to upload an allele file, if the user has one.

A single peptide length, multiple peptide lengths, or a range of peptide lengths can be chosen under “Select length(s)”. If the user has input individual peptide sequences into the tool (as is the case here) the lengths selected should correspond with the lengths of these sequences. Otherwise, if the user wants to determine epitopes of various lengths in a non-specific manner,

they may do so, though this may be ineffective without inputting either a full protein sequence or large peptide sequences. There is an option for the user to “Sort peptides” by either “Adjusted Rank” or “Position in Sequence”, though “Adjusted Rank” makes it easier to filter potential epitopes from other peptides in the results. “Output format” must be specified as either an “XHTML table” or a “Text file” and an email address can be entered in order for users to receive their results if the server times out.

Note: This manual will only discuss human allele options, though it is possible to indicate mouse alleles (H-2-I) using the MHC II binding prediction tool.

Step 3: The results page specifies the allele and potential epitope sequence of each binding prediction. IEDB recommends that users use a 20 percent and higher adjusted rank threshold in order to screen the resulting peptides for epitopes. The user should only consider these peptides as potential epitopes. The method used for each peptide-MHC II binding prediction is specified, as are the start and end positions of each peptide (on the protein or peptide sequences inputted by the user).

Note: For a peptide that has an adjusted rank of 20.0, this means that the peptide’s binding score is ranked 20th percentile out of a large set of randomly selected peptides of similar length sourced from the SWISS-PROT database. Binding scores are IC50 values and smaller values indicate higher binding affinity. Therefore, the smaller the adjusted rank, the better the binding affinity of the peptide. Adjusted rank is merely the “percentile rank” adjusted to remove bias resulting when there are multiple peptide lengths involved (adjusted rank takes into account the frequency of each peptide length in adjusting percentile rank).

Note II: The recommended adjusted rank threshold for MHC I binding predictions is 1 percent.

3. Finding novel MHC Class II epitopes of an antigen using binding affinity and proteosomal cleavage prediction tools (*not included in video tutorial*)

Step 1: Find the full protein sequence of the antigen of interest. From the IEDB homepage, navigate again to the MHC Class II binding prediction tool. Copy it to the clipboard.

Step 2: Paste it in the sequence input field. User should choose their prediction method, alleles they intend to explore, and the lengths. Results should be sorted by adjusted peptide rank. User should also include their email in order to receive a .csv file they can open in Excel to analyze results.

Step 3: Once results are in and Excel file is open, user should navigate back to the homepage and select “MHC I Processing” under “T Cell Epitope Prediction”. In the new page, user should select “MHCII-NP” under “T Cell Epitopes – Processing Prediction”.

Step 4: The tool has a single field for entering a protein sequence. User should again copy the protein sequence of their antigen of interest and paste it in the sequence field, then click “Submit”. The results will list the top 5 peptides, sorted by cleavage probability percentile rank.

Note: As with “adjusted ranking” in the MHC Class II binding prediction tool, the lower the “cleavage probability percentile rank” the greater the likelihood that the peptide is an MHC II ligand.

Step 5: Copy the first peptide in the results. Returning to the excel file, paste the peptide sequence into the Ctrl-F search field, then hit “find next”. Verify whether the adjusted ranking falls below the 20.0 threshold. If it does, note this peptide. If not, disregard it. Repeat this step with each of the other top 5 peptides in the processing results. You have now used two tools to quickly filter your antigen of interest for peptides that are most likely to be epitopes.

Video Tutorial

Link: https://www.youtube.com/watch?v=JH9pwsXDUNk

References

Lundegaard, C., Lund, O., Buus, S., & Nielsen, M. (2010). Major histocompatibility complex class I binding predictions as a tool in epitope discovery. *Immunology*, *130*(3), 309–318. doi: 10.1111/j.1365-2567.2010.03300.x

Flower, D. R., & Perrie, Y. (2013). Immunomic Discovery of Adjuvants and Candidate Subunit Vaccines. New York, NY: Springer. doi: https://doi.org/10.1007/978-1-4614-5070-2

Bardi, M. S., Jarduli, L. R., Jorge, A. J., Camargo, R. B. O. G., Carneiro, F. P., Gelinski, J. R., … Lavado, E. L. (2011). HLA-A, B and DRB1 allele and haplotype frequencies in volunteer bone marrow donors from the north of Parana State. *Revista Brasileira De Hematologia e Hemoterapia*, *34*(1), 25–30. doi: 10.5581/1516-8484.20120010

Stylianou, E., Harrington-Kandt, R., Beglov, J., Bull, N., Pinpathomrat, N., Swarbrick, G. M., … McShane, H. (2018). Identification and Evaluation of Novel Protective Antigens for the Development of a Candidate Tuberculosis Subunit Vaccine. Infection and Immunity, 86(7), 1–16. doi: 10.1128/IAI.00014-18

Paul, S., Arlehamn, C. S. L., Scriba, T. J., Dillon, M. B., Oseroff, C., Hinz, D., … Sette, A. (2015). Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *Journal of Immunological Methods*, *422*, 28–34. doi: 10.1016/j.jim.2015.03.022

Martini, S., Nielsen, M., Peters, B., & Sette, A. (2019). The Immune Epitope Database and Analysis Resource Program 2003–2018: reflections and outlook. *Immunogenetics*, 1–20. doi: 10.1007/s00251-019-01137-6

Cruz-Tapias P, Castiblanco J, Anaya JM. Major histocompatibility complex: Antigen processing and presentation. In: Anaya JM, Shoenfeld Y, Rojas-Villarraga A, et al., editors. Autoimmunity: From Bench to Bedside [Internet]. Bogota (Colombia): El Rosario University Press; 2013 Jul 18. Chapter 10. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459467/

Annaya, J.-M., Shoenfeld, Y., Rojas-Villarraga, A., Levy, R. A., & Cervera, R. (2013). *Autoimmunity: From Bench to Bedside*. Bogota, Colombia: El Rosario Free Press.