A new version of the Analysis Resource was deployed on June 24 to update the MHC class II epitope prediction methods. Two new methods were introduced – a combinatorial peptide scanning library and NN-align. NN-align is an artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. It simultaneously identifies the MHC class II binding core and binding affinity. The method is trained using an algorithm that corrects bias in the training data caused by redundancy in binding core representation. Prediction accuracy has been shown to improve significantly when information about the residues flanking the peptide-binding core is taken into account. A 2009 paper in BMC Bioinformatics by Nielsen and Lund describe the method in detail (PMID: 19765293).

With the introduction of two new class II epitope prediction methods, bioinformatics specialists at LIAI updated the consensus prediction method to include them. The new consensus method uses NN-align, SMM-align, and the combinatorial peptide scanning library. When the scanning library is not available for an allele, the Sturniolo method is used instead. In addition, a new MHC class II binding dataset was made available for download from the website (http://tools.immuneepitope.org/analyze/html/download_MHC_II.html). It contains over 44,000 experimentally measured binding affinities covering 26 allelic variants. The dataset can be downloaded in a tar.gz format.

The latest version of the IEDB website was released in July. Version 2.4 added several new features that improve usability. The allergen tree in the Source Organism Finder was updated to display only nodes referenced by curated data, thus eliminating searches that return no values. Other changes affect the home page search feature. The user can now specify epitope structure in their search as linear peptide, discontinuous structure, non-peptidic, or all of the above (“any”) by selecting among the four radio buttons. An autocomplete feature

Continued on Page 2
Continued from Page 1

has also been implemented the Molecule, Organism, and Allele Finders. For example, as one types “hem” in the Source Antigen field in the Epitope Source section, several choices start to appear in a list below the text field, including “hemolysin”, “hemK protein”, and “hemagglutinin”. The user can select from the list using a mouse or can continue typing to further narrow down choices. Likewise, a user may type “human” or “homo” to select “homo sapiens” in the Host Organism field. The list of matches includes scientific names and synonyms. These actions enable users to quickly specify their search parameters without using the finders. The other significant feature included in IEDB 2.4 is the new Molecule Finder described in another article in this newsletter.

Non-Peptidic and Protein Trees Introduced into Molecule Finder

Over a year ago it became apparent to the IEDB team that the source molecule finder did not fully meet the needs of the IEDB user base. In order to select a specific source molecule, the user needed to type its name in a free text field. Although this was a workable method, it proved to be error-prone and assumed that the user was familiar with the contents of the database. Two of the most requested features of our users were: 1) the ability to browse the available source molecules in a meaningful way and 2) the ability to select a source protein and have all related proteins added to the query as well. To this end, the source molecule finder was redesigned to include two parallel trees, one for non-peptidics and the other for protein molecules. The former contains the structures curated by the Chemical Entities of Biological Interest (ChEBI) database. An example is shown in Figure 1.

The development team determined that the most logical way to group the proteins was by organism. In order to accomplish this, the NCBI species was determined for each of the proteins in the database. For viruses and bacteria, this involved traversing the NCBI taxonomy from the sub-species (strain) level up to the species level. For each

Continued on Page 3
species, a set of reference proteins was selected from the NCBI protein database based upon the availability of a complete genome for the species. All proteins for each species were BLASTed against the reference protein set to determine their homologs. These data were used to build the protein tree in a way that mirrors a pruned version of the NCBI taxonomy. The result is a coherent tree that is divided along major taxonomic categories and is quickly traversed with proteins grouped logically below each species. The ability to perform a free text search was maintained and is now enhanced by the ability to see where the protein is placed in the tree. It is now possible, for example, to select all Influenza A haemagglutinin (HA) proteins by selecting one node of the tree rather than individually clicking on the 100+ different HA proteins in the database. This can be seen in Figure 2.

Figure 2 Example of the protein tree as found in the Source Antigen Molecule Finder on the IEDB home page
Meta-analysis of all immune epitope data in the Flavivirus genus: inventory of current immune epitope data status in the context of virus immunity and immunopathology.

Vaughan K, Greenbaum J, Blythe M, Peters B, Sette A.


PMID: 20565291

A meta-analysis was performed in order to inventory the immune epitope data related to viruses in the genus Flavivirus. Nearly 2000 epitopes were captured from over 130 individual Flavivirus-related references identified from PubMed and reported as of September 2009. This report includes all epitope structures and associated immune reactivity from the past and current literature, including: the epitope distribution among pathogens and related strains, the epitope distribution among different pathogen antigens, the number of epitopes defined in human and animal models of disease, the relationship between epitopes identified in different disease states following natural (or experimental) infection, and data from studies focused on candidate vaccines. We found that the majority of epitopes were defined for dengue virus (DENV) and West Nile virus (WNV). The prominence of DENV and WNV data in the epitope literature is likely a reflection of their overall worldwide impact on human disease, and the lack of vaccines. Conversely, the relatively smaller number of epitopes defined for the other viruses within the genus (yellow fever and Japanese encephalitis virus) most likely reflects the presence of established prophylaxis and/or their more modest impact on morbidity and mortality globally. Through this work we hope to provide useful data to those working in the area of Flavivirus research.

IEDB at FOCIS Conference

The IEDB exhibit booth was present at the FOCIS (Federation of Clinical Immunology Societies) 2010 conference in Boston on June 24 – 27. The booth was staffed by IEDB senior curator Ken Chan, PhD and Data Quality Control and Assurance Manager Randi Vita, MD. The exhibits were available Thursday and Friday evenings. Approximately 40 scientists visited with questions regarding the content and usage of the IEDB. Many other participants stopped by to pick up literature describing the database. Scientists with interest in the IEDB worked both in academics and industry and were largely from the United States, but several worked in Canada, the United Kingdom, Germany, India, or other countries. At this conference interest in the IEDB was equal among PhD and MD researchers. Of those visitors who had previously used the IEDB, all had positive feedback to give. Many had not previously used the database, but had heard of it and were interested in obtaining more information. Several scientists had questions regarding epitopes from cell surface receptors and monoclonal antibodies.
The first half of 2010 has been a productive time for data submissions to the IEDB from researchers. A total of 5,227 epitopes have been submitted and most of them are visible to the scientific community. Submitters have the option of putting their data on hold until they have been published. All were submitted using the IEDB’s Data Submission Tool (DST). The DST is designed to facilitate the process of data submission to the IEDB. As the first step, the submitter is provided with tab-delimited template files corresponding to the various types of data that can be submitted (MHC binding, MHC Ligand Elution, T cell, B cell), as well as a submission file where important submission details are reported. These files serve as data entry forms where the submitter formats the data according to guidelines. Upon completion, the submitter then submits the validated files to the IEDB, where IEDB personnel review and transfer the data to the IEDB website where it will be publicly accessible. Three T cell data templates were recently redesigned to accommodate data from the allergy epitope discovery contracts. Each has been successfully utilized to submit data to the IEDB.

Curation of data relating to peptidic epitopes for all infectious diseases and peptidic and non-peptidic epitopes for allergens is current for references appearing in PubMed as of the end of March 2010. A query for new potentially relevant epitope references is run quarterly to update the database. Curation of peptidic epitopes for diabetes, rheumatoid arthritis, multiple sclerosis, and lupus is essentially complete. Curation of peptidic epitopes of all other autoimmune diseases and non-peptidic epitopes for all infectious diseases is in progress and will be completed later in 2010. As of August 2010, data from approximately 10,600 references have been incorporated into the IEDB. The IEDB contains data for over 76,000 epitopes, 2,500 epitope source organisms, and 554 restricting MHC alleles. Users are invited to bring references to our attention that are potentially relevant to the IEDB but do not appear in the database. References that are deemed to meet the IEDB criteria for curation will be queued for processing in accordance to our NIAID-directed priorities (Category A-C priority pathogens, emerging and re-emerging infectious diseases, other infectious diseases, allergies, autoimmune diseases, and transplantation). Citations should be sent to help@iedb.org.

The Immune Epitope Database and Analysis Resource is supported by a contract from the National Institute of Allergy and Infectious Disease, NIH, DHHS (Contract HHSN266200400006C). The newsletter is distributed four times a year. We welcome communication from the users of the IEDB database and invite suggestions for articles in future issues. To subscribe to the IEDB newsletter or to contact project staff, send your email information to the email address below.

Email: contact@iedb.org
Web: http://www.iedb.org

Principal Investigator:
Alessandro Sette, Ph.D.
alex@liai.org

Co-Principal Investigator:
Bjoern Peters, Ph.D.
bpeters@liai.org

Immune Epitope Database and Analysis Resource
c/o La Jolla Institute for Allergy & Immunology
9420 Athena Circle
La Jolla, CA 92037
(858) 752-6500

Project Director:
Stephen Wilson, Ph.D.
swilson@liai.org

Production:
Emily Seymour
Ward Fleri, Ph.D.