

# Immune Epitope Database NEWSLETTER

Volume 2, Issue 2

<http://www.immuneepitope.org>

July 2005

## A Focus on the Curation Process

A great deal of effort has been expended the last several months in literature curation. We have significantly increased the number and level of expertise of our curation staff, improved quality, and refined processes. La Jolla Institute for Allergy & Immunology (LIAI) now employs five full-time curators and three part-time curators whose sole function is to read and curate epitope information found in the scientific literature. Other IEDB team members play significant roles in training curators and reading article abstracts to decide whether the article has content that should be curated, as well as curating several articles each week. Epitope Council (EC) members, with their immunological expertise and knowledge of the Immune Epitope Database (IEDB) data structure, have been instrumental in assuring the quality of the curators' output. The experience gained in curation has been critical in providing the database development team with new insights for improving the data structure and the curation process itself.



Weekly curation meetings are held in the Bunker Hill conference room with the full curation staff to go over questions, solutions, new ideas, and anything else that will aid in improving the process.

## The Steps to the Curation Process & Work Flow Chart

The curation team is currently focused on curating references that discuss epitopes from the NIAID Category A-C priority pathogens (website: [http://www2.niaid.nih.gov/Biodefense/bandc\\_priority.htm](http://www2.niaid.nih.gov/Biodefense/bandc_priority.htm)), starting with the most recent articles and moving backwards in time. The steps of the curation process are described below and are displayed graphically in the accompanying figure on page 4.

### *Identifying References*

For literature curation, references are identified using PubMed queries. The query to retrieve all category A-C priority pathogen related manuscripts yielded 3530 references from PubMed.

continued on page 4

## Inside This Issue

- 1.) A Focus on the Curation Process  
Ramping up Literature Curation
- 2.) The Steps to the Curation Process  
& Workflow Chart
- 3.) Meet Our New Curators  
Featuring 4 New Faces
- 4.) Recent Publications
- 5.) Visitations to the Lab

# Four New Curators

*featuring Nima, Russell, Jong & Eugene*

---

## *Nima Salimi*

Nima went to undergraduate school at the Univ. of Calif., San Diego and received a B.S. in Physiology and Neuroscience. He then went on to Georgetown University to obtain a M.S. in Physiology and Biophysics.

He read about the awarding of the NIH grant to LIAI in the San Diego Union Tribune and hope to expand his knowledge of immunology. He would like to apply this knowledge to help make the IEDB a more powerful tool for other scientists.

He feels that while traditional research offers an in-depth exploration of a few, very specifically focused realms of science, being a curator affords the daily opportunity to visit diversified research endeavors worldwide. He can learn about malaria studies in South America in one hour, while learning about progress toward a SARS vaccine in the next hour. However, staring at the computer screen for eight hours per day can be a little hard on the eyes.

He was a spinal cord injury researcher at UCSD Department of Neurosciences. Please see <http://www.pnas.org/cgi/reprint/98/6/3513> for publication.



## *Eugene Moore*

Eugene went to the University of Texas and the University of Arkansas for his undergraduate studies where he received a B.S. in Microbiology. He later attended the Univ. of Calif., Riverside to obtain a Ph.D. in Biochemistry. His doctoral and post-doctoral work focused on the biochemical identification and characterization of novel insect toxins utilizing bioanalytical techniques such as HPLC and mass-spectrometry (MS and MS/MS), molecular biology, protein chemistry, and electrophysiology.



As a curator, he is building upon his knowledge base on current trends and methodologies in his areas of interest while coincidentally providing an invaluable source of data for practicing scientists who are active in vaccine discovery projects and pharmaceutical drug development. Ultimately, he would like to learn how to develop the most practical bioinformatics database possible from the standpoint of data entry and from the end-user perspective of tools development and query features.

His interest in science comes from a natural interest in how life works on the micro levels from the biophysical, biochemical, and neuroethology perspectives to the macro levels of social and evolutionary biology. Curation therefore serves as one of my primary driving forces of learning, compiling, and putting the pieces together into an integral construct of the mechanics of life. He wishes to be a strong contributing factor in the initial and on-going stages of the database development to make this project as successful as possible.

## *Russell Chan*



Russell went to undergraduate school at Brown University and received an A. B. in Biology. He then went on to M. I. T. to obtain a Ph.D. in Microbiology. He did his postdoc training at the University of Washington in the Genetics Dept.

He read about the job opportunity on the LIAI website and hopes to achieve an in-depth understanding and appreciation of current developments in immunology from working on the IEDB.

Being a curator provides the satisfaction of knowing that this project can have a real impact on medical research that can improve human health and our quality of life.

Prior to working on the IEDB, he was a research Scientist in the Forage Additives Research group at Pioneer Hi-Bred International, Inc., Johnston, IA, where he created and managed the database infrastructure (including over 15 Microsoft Access databases) for a Strain Collection of 20,000 microbial isolates. He also held positions with Biogen (Cambridge, MA and Geneva, Switzerland) and on the faculty of the University of Cincinnati College of Medicine. His most memorable scientific accomplishment is when he discovered the transposon Tn10 in the Salmonella phage P22 and the yeast SST2 gene (Supersensitive to alpha factor pheromone) that was subsequently recognized as the first member of a superfamily of evolutionarily conserved proteins, called RGS proteins, for “regulators of G protein signaling”.

## *Jong De Castro*



Jong went to undergraduate school at the University of the Philippines and received a B.S. in Molecular Biology & Biotechnology. He then attended the University of Texas Medical Branch where he obtained a Ph.D. in Human Biological Chemistry & Genetics.

Immunology, particularly in vaccine design has always been an interest of his. His undergraduate thesis topic was on designing synthetic peptide vaccines. He feels this is his last chance to learn about how people do it these days. He was doing it 15 years ago.

He has never read so many interesting papers in a very short period of time (except when he was writing his dissertation). He's also learning a lot about putting together a scientific database.

Before becoming a curator on the IEDB, he was a bench researcher and a teacher. He held a couple of postdoctoral positions in the neurosciences and taught General Biology at the San Diego City College.

# Recent Publications

**Huynh-Hoa Bui, Ph.D. - Immunogenetics. 2005 Jun;57(5):304-14.**

**Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications.**

Bui HH, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton KA, Mothe BR, Chisari FV, Watkins DI, Sette A.

Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, 3030 Bunker Hill Street, Suite 326, San Diego, CA 92109, USA.

Prediction of which peptides can bind major histocompatibility complex (MHC) molecules is commonly used to assist in the identification of T cell epitopes. However, because of the large numbers of different MHC molecules of interest, each associated with different predictive tools, tool generation and evaluation can be a very resource intensive task. A methodology commonly used to predict MHC binding affinity is the matrix or linear coefficients method. Herein, we described Average Relative Binding (ARB) matrix methods that directly predict IC50 values allowing combination of searches involving different peptide sizes and alleles into a single global prediction. A computer program was developed to automate the generation and evaluation of ARB predictive tools. Using an in-house MHC binding database, we generated a total of 85 and 13 MHC class I and class II matrices, respectively. Results from the automated evaluation of tool efficiency are presented. We anticipate that this automation framework will be generally applicable to the generation and evaluation of large numbers of MHC predictive methods and tools, and will be of value to centralize and rationalize the process of evaluation of MHC predictions. MHC binding predictions based on ARB matrices are made available at <http://epitope.liai.org:8080/matrix> web server.

PMID: 15868141 [PubMed - in process]

---

**Bjoern Peters, Ph.D. - Cell Mol Life Sci. 2005 May;62(9):1025-37.**

**Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method.**

In an effort to document algorithms that will be utilized within epitope identification tools of the analysis resource section of the IEDB, a manuscript was published in the May issue of BMC Bioinformatics describing the stabilized matrix method (SMM) [1]. This is one of several existing methods used to generate matrices quantifying the influence of each possible amino acid in a peptide on the efficiency of a biological process. This method has been applied to predict peptide binding to MHC [2], transport of peptides into the ER by TAP [3] and cleavage of proteins by the proteasome [4]. The current publication describes the applied algorithm in detail, and makes a software package implementing it available to the public. This is especially important to ensure that tool generation and evaluation within the IEDB project can be reproduced by interested members of the scientific community.

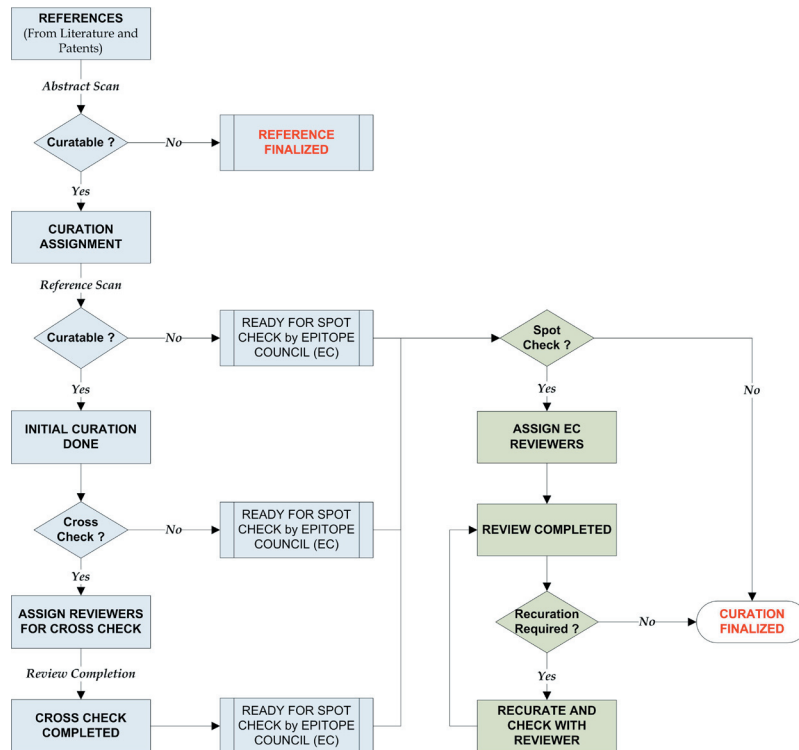
[1] Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics. 2005 May 31;6(1):132.

[2] Peters B, Tong W, Sidney J, Sette A, Weng Z. Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules. Bioinformatics. 2003 Sep 22;19(14):1765-72.

[3] 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.

[4] 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.

# Curation Process Flow Chart



continued from page 1

## *Abstract Scan*

The Abstract Scan is the first step to identify references that contains curatable data. In Abstract Scan, the abstract of the reference is read or “scanned” by an EC member. Approximately 1250 of the 3530 abstracts were reviewed during the reporting period. Of this number, 430 references (approximately 35%) were judged to be curatable.

## *Initial Curation (Reference Scan)*

References identified as curatable after the abstract scan undergo initial curation. Some references may be deemed uncuratable during this step after the entire paper has been read. After initial curation, references proceed to either the cross-check or spot-check step.

## *Cross Check*

References sent for cross check are reviewed by second-tier curators (curation supervisors). This step is typically performed when new curators are being trained. The reviewer addresses any curation problems or errors directly with the original curator. The curation assessment and other date/time information is recorded in the curation tracking system. Once a curated reference is finalized by the second-tier curator, the reference proceeds to the spot-check step.

## *Spot Check*

All the references that are finalized by the curators are available for Spot Check. Spot checks are performed by Epitope Council (EC) members. Currently all the curated references are being reviewed by the EC members. As the curation team gains experience, we anticipate that the need to check all curated references will decrease and we will move to actual spot checks.

continued on page 5

## Visitations

### Darren R. Flower

Edward Jenner Institute for Vaccine Research  
Head of Bioinformatics Group

May 24, 2005

“Computational Vaccinology at the EJIVR”

May 25, 2005

“The AntiJen Database”

### Andrey Rzhetsky

Columbia University  
Department of Bioinformatics

August 24, 2005

“A Bird’s Eye View of Text Mining for Biology”

continued from page 4

In order to perform curation before the IEDB production system is completed, LIAI has been using an Interim Curation System (ICS), built by team member SAIC, and a Curation Tracking System (CTS), developed by LIAI. Both systems use Microsoft Access. The ICS uses the same database schema as the production system currently under development. All curated data in the ICS will be migrated to the IEDB prior to beta testing, which is scheduled to begin on November 1.

The CTS tracks all the process involved in finalizing a reference. References that come out as hits from querying bibliographic interfaces (PubMed, WIPO) are recorded in the tracking system. Only those references that are marked curatable after abstract scan are available for curation assignment. Parameters such as date of curation assignment, time take to complete the initial curation, date of completion, and number of epitopes discussed in the reference are stored in the tracking system. After curation review, various parameters like date/time information and assessment about curation are stored in the system.

To standardize the curation process, the curation team has been recording notes of standards and procedures, which has evolved into a curation manual. This document continues to grow as new methods and concepts are encountered in the literature. The manual has been exceedingly useful for training new curators and developing uniformity among the curators and the reviewers.

## Contact Information

The Immune Epitope Database is supported by a contract from the National Institute of Allergy & 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. Upon deployment of the database, we will actively solicit tool and epitope submissions. To subscribe to the IEDB newsletter or contact project staff, send your email information to the email address below.

Immune Epitope Database and Analysis Resource  
c/o La Jolla Institute for Allergy & Immunology  
s10355 Science Center Drive  
La Jolla, CA 92121

(858) 558-3500

Email: [contact@immuneepitope.org](mailto:contact@immuneepitope.org)

### Principal Investigator:

Alessandro Sette, Ph.D. - [alex@liai.org](mailto:alex@liai.org)

### Project Director:

Stephen Wilson, Ph.D. - [swilson@liai.org](mailto:swilson@liai.org)

### Contributors:

Muthu Sathiamurthy

### Writers:

Ward Fleri, Ph.D.

### Production:

Jody Dang  
Michael Scarpelli