

Bjoern Peters · John Sidney · Phil Bourne · Huynh-Hoa Bui · Soeren Buus · Grace Doh · Ward Fleri · Mitch Kronenberg · Ralph Kubo · Ole Lund · David Nemazee · Julia V. Ponomarenko · Muthu Sathiamurthy · Stephen P. Schoenberger · Scott Stewart · Pamela Surko · Scott Way Steve Wilson · Alessandro Sette

The design and implementation of the immune epitope database and analysis resource

Received: 22 December 2004 / Revised: 22 April 2005 / Published online: 14 May 2005
© Springer-Verlag 2005

Abstract Epitopes are defined as parts of antigens interacting with receptors of the immune system. Knowledge about their intrinsic structure and how they affect the immune response is required to continue development of techniques that detect, monitor, and fight diseases. Their scientific importance is reflected in the vast amount of epitope-related information gathered, ranging from interactions between epitopes and major histocompatibility complex molecules determined by X-ray crystallography

to clinical studies analyzing correlates of protection for epitope based vaccines. Our goal is to provide a central resource capable of capturing this information, allowing users to access and connect realms of knowledge that are currently separated and difficult to access. Here, we portray a new initiative, “The Immune Epitope Database and Analysis Resource.” We describe how we plan to capture, structure, and store this information, what query interfaces we will make available to the public, and what additional predictive and analytical tools we will provide.

B. Peters · J. Sidney · H.-H. Bui · W. Fleri · M. Kronenberg · R. Kubo · M. Sathiamurthy · S. P. Schoenberger · S. Wilson · A. Sette (✉)
La Jolla Institute of Allergy and Immunology,
3030 Bunker Hill Street, Suite 326,
San Diego, CA, 92109, USA
e-mail: alex@liai.org
Tel.: +1-858-2281362
Fax: +1-858-2281008

P. Bourne · J. V. Ponomarenko
San Diego Supercomputer Center,
P.O.Box 85608 San Diego, CA, 92186-5608, USA

S. Buus
University of Copenhagen,
Panum Building 18.3.22, Blegdamsvej 3,
220 Copenhagen, Denmark

G. Doh
SH Grace Consulting,
A-402 Hannam Riverhill, 390 Hannam-dong, Youngsan-ku,
Seoul, 140-210, South Korea

O. Lund
BioCentrum-DTU, Technical University of Denmark,
Building 208, Lyngby, 2800, Denmark

D. Nemazee
Department of Immunology, The Scripps Research Institute,
10550 North Torrey Pines Road, IMM-29,
La Jolla, CA, 92037, USA

S. Stewart P. Surko S. Way
Science Applications International Corporation,
9455 Towne Center Drive, MS-W2,
San Diego, CA, 92121, USA

Keywords Epitope · Antibody · MHC · Database

Introduction

Over the last 20 years, the amount of information related to epitopes recognized in the course of T- and B-cell-mediated immune responses has dramatically increased. As of November 2004, a PubMed search using the word “epitope” revealed a total of 5,173 records prior to 1975, and 17,088 records in the 1975–1984 period. The number of records jumps to 30,948 for 1985–1994, and has now reached 34,205 for the 1995–2004 period.

Epitope-based techniques allow the accurate and precise characterization of immune responses. This precise definition of immune responses allows understanding and quantitating instances of immune responses in which the microbe is able to overcome the host, or establish chronic infection. This knowledge is ultimately crucial for basic immunological studies, and for designing effective intervention strategies and vaccines, as well as for the rigorous evaluation of new vaccine candidates. Knowledge of the epitopes recognized in the course of natural infection, or as a result of vaccination, is also key for our capacity to develop bioinformatics tools and to accurately analyze, model, and predict immune responses. Recently, a renewed focus on emerging diseases and bioterrorism has emphasized the need to develop a central repository of available information regarding the key epitopes recognized by the immune system.

Despite this growing need, a centralized resource to store and access relevant information is not available. To address this need, the National Institute for Allergy and Infectious Diseases (NIAID) is sponsoring the creation of the Immune Epitope Database and Analysis Resource (IEDB), a comprehensive knowledge center composed of a repository of immune epitope data and associated analysis tools. This repository will host data relating to both B and T cell epitopes from infectious pathogens as well as experimental and self-antigens (RFP-NIH-NIAID-DAIT-03/31). Priority will be placed on pathogens considered to be potential bioterrorism threats and emerging diseases. Epitopes recognized in humans, nonhuman primates, rodents, and other animal species are included. In a coordinated effort to increase our knowledge in the field of emerging diseases and category A–C pathogens, NIAID has also launched a large-scale antibody and T cell epitope discovery program aimed at generating epitope data and analysis resources to be included in the IEDB (RFP-NIAID-DAIT-03-43/-04-39).

The IEDB is not the first database aimed at storing immune epitope information. Table 1 gives an overview of existing databases. By examining their structure during the IEDB design, we were able to capitalize on their experiences. Most components of the IEDB can be found in one of these databases, but none contains them all. For example, the pioneering SYFPEITHI database focuses on carefully mapped epitopes or naturally processed peptides, but does not further annotate the context in which an epitope is immunogenic. The Los Alamos HIV Molecular Immunology Database does provide an extensive context annotation, but is limited to a single virus system.

So what are the main distinguishing features of the new IEDB database? First, we will place a priority focus on category A–C pathogens. Second, database dimensions compatible with inclusion of a large number of records will be provided, which is needed to accommodate further growth due to large-scale epitope identification programs. Third, the IEDB will include T cell and B cell epitope data from both humans and other animal species. Fourth, extensive epitope annotation will allow users to customize queries to an unprecedented level of detail (e.g., search by type of assay, type of cytokine released in response, and route of vaccine administration). Finally, the inclusion of

all available data including patent data will provide unprecedented comprehensiveness. The patent literature is underrepresented by most epitope databases, but contains a large fraction of the data generated by the biotechnology and pharmaceutical industries.

Another crucial component of the project is the involvement of the scientific community in the design of the IEDB's scope and capability. The IEDB will be produced in a manner that encourages the incorporation of data and analytical tools derived by research labs at large. The scope of this report is to inform the scientific community of our effort and to solicit feedback while the project is still in a design stage. Herein, we present the specific immunological perspective of the IEDB and the database concepts applied to its design. We envision that this resource center will be freely available on the Internet, and that a prototype will be available and operational in the first quarter of 2006.

Results

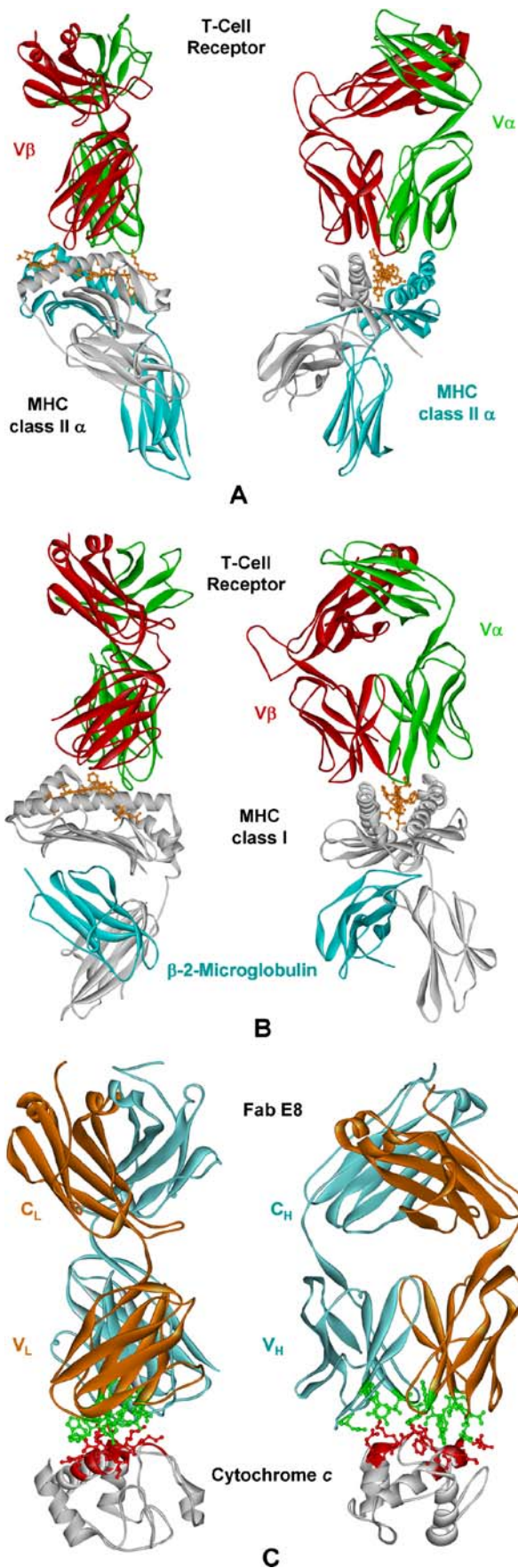
What is an epitope? Defining the IEDB scope

An immunogen is defined as a substance used to elicit a specific immune response, whereas an antigen is a substance recognized by an existing immune response and associated molecules such as T cells or antibodies in a recall response. An epitope is defined as the chemical structure recognized by antigen specific receptors of the immune system [antibodies and/or T cell receptors (TR)] (William 1999). In the case of most T cell epitopes, an epitope is defined as a structure that is presented in association with a specific MHC, and is bound by the variable domain of a specific TR (Fig. 1a, b). Likewise, a B cell epitope is defined as a structure bound by the variable domains of a monoclonal or polyclonal antibody (Fig. 1c). Both linear and conformational B cell epitopes are considered within the scope of our project. Linear epitopes (also called sequential or continuous epitopes) are normally defined as short peptide fragments that can bind antibodies raised against the intact protein from which the peptides are derived. Conformational or discontinuous epitopes are defined as the set of antibody binding amino acids that are not adjacent in the primary sequence, but are brought into proximity by the folding of the polypeptide chain. The classification is not clear-cut as discontinuous epitopes may contain linear stretches of amino acids, and continuous epitopes may contain indifferent amino acids. Ninety percent of antibodies raised against proteins react against discontinuous fragments (Van Regenmortel 1996).

Over the course of the last 20 years, a sizable amount of data has been accumulated relating to which peptides can bind to specific MHC (Barber and Parham 1993; Buus et al. 1987; Engelhard 1994; Engelhard et al. 2002; Fleckenstein et al. 1999; Flower et al. 2003; Gromme and Neefjes 2002; Hammer et al. 1993; Lauemoller et al. 2000; Lehner 2003; Madden 1995; Margalit and Altuvia 2003; Rammensee 1995; Saveanu et al. 2002; Sette and Sidney

Table 1 Existing epitope databases

Name	Link
ΦIMM	http://www.sdmc.i2r.a-star.edu.sg:8080/fimm/
HLA Ligand/ Motif database	http://www.hlaligand.ouhsc.edu/
JenPep	http://www.jenner.ac.uk/JenPep/
Los Alamos HIV Molecular Immunology database	http://www.hiv-web.lanl.gov/content/immunology/index.html
MHCBN	http://www.imtech.res.in/raghava/mhcbn
SYFPEITHI	http://www.syfpeithi.de



◀ **Fig. 1** Orthogonal views of X-ray structures of immunological complexes available in the Protein Data Bank. **a** Complex of a human TR, influenza Ha antigen peptide PKYVKQNTLKLAT, and MHC Class II HLA-DR1 (PDB ID, 1fyt). T-cell epitope residues colored in orange. **b** Xenoreactive complex AHIII 12.2 TR bound to P1049 ALWGFFPVLs and MHC Class I HLA-A2.1 (PDB ID, 1lp9). T-cell epitope residues colored in orange. **c** IgG1 Fab fragment of E8 antibody complexed with horse cytochrome *c* (PDB ID, 1wej). B cell epitope and paratope residues colored in red and green, accordingly. The figures were obtained using the WebLab viewer

1999; Stevanovic 2002; van der Merwe and Davis 2003; Van Kaer 2002). In addition, a wealth of data has been generated relating to the peptides naturally presented by MHC (Engelhard et al. 2002; Falk et al. 1991; Hickman et al. 2004; Latek and Unanue 1999; Rammensee et al. 1993; Rotzschke et al. 1990; Shastri et al. 1998; Stevanovic et al. 2003; Van Bleek and Nathenson 1990). Taken together, these data have provided the basis for clarification of the chemical specificity of MHC, and for the development of methods to predict and identify peptides binding to, or naturally presented by, MHC (Rotzschke et al. 1991). The fact that a given structure can bind to or is presented by MHC does not guarantee, per se, that the resulting complex will be capable of engaging specific TR and thereby meet the definition of an epitope. However, for the sake of completeness, the IEDB will also include data relating to MHC binding, or lack thereof, and naturally presented MHC ligands since compounds capable of binding to MHC have the potential of being immunogenic or antigenic for T cells.

Intrinsic and extrinsic features associated with epitopes

In designing the database, we started from the basic observation that each epitope is associated with intrinsic and extrinsic features. Intrinsic features are those determined by its structure and its phylogenetic relations, whereas extrinsic features are context-dependent attributes dependent on the experimental or natural environment. In other words, immune epitope information is best represented by capturing not only genomic/proteomic and structural information regarding the pathogen of origin, but also representing the immune system of the host interacting with the pathogen, and the biological circumstances in which the interaction occurs. Accordingly, in our view, information related to epitopes should be represented by considering both intrinsic and extrinsic features.

At the level of T cell epitopes, intrinsic features that we plan to include in the IEDB are the molecular structure of the epitope, its binding affinity for different MHC, and the affinity of epitope/MHC complexes for TR of defined sequence. Likewise, at the level of B cell epitopes, we plan to include intrinsic features such as the epitope's molecular structure and binding affinity for immunoglobulin (IG) or antibody of defined sequence. These features are unequivocally specified and can be singularly associated with a given epitope structure or receptor/epitope combination. Finally, each epitope may also be linked with objective and

quantifiable measurements of its binding affinity to specific immune receptors.

As mentioned above, other features, such as immunogenicity, are not intrinsically associated with a given epitope or structure, but rather are highly context-dependent (i.e., extrinsic). Accordingly, lack of information regarding the context may limit the usefulness of data relating to the epitope's antigenicity or immunogenicity. Context information includes data representing the immune system of the host interacting with the pathogen and the biological circumstances in which the interaction occurs. Examples of context information and extrinsic features are the species of the host (or vaccine recipient) in which the response occurs, the assay utilized to measure responses, and the dose and route of administration (Brockstedt et al. 1999). Likewise, the yield when processing a given epitope protein precursor is dictated by the overall sequence of the protein in which the epitope is contained, and the general cellular environment in which cellular processing occurs (York et al. 1999).

3D structure as an intrinsic feature associated with epitopes

The Protein Data Bank (Berman et al. 2000a,b) currently provides access to 3D structures of more than 1,600 molecules of immunological interest and their complexes. These structures are valuable for epitope analysis and can be used for prediction tool development. For these purposes, as well as for epitope visualization, the IEDB will provide access to the available 3D structures of epitope/MHC and TR/epitope/MHC complexes, antibodies and antibody/antigen complexes, and proteins of the immunoglobulin superfamily (IgSF) and MHC superfamily (MhcSF) in the PDB. Figure 1 presents examples of X-ray structures of epitopes in complexes with antibody and TR/MHC: (1) complex of a human TR, Influenza Ha antigen peptide, and MHC Class II HLA-DR1 (Hennecke et al. 2000); (2) xenoreactive complex AHIII 12.2 TR bound to P1049 ALWGFFPVLS and MHC Class I HLA-A2.1 (Buslepp et al. 2003); (3) E8 antibody Fab complexed with horse cytochrome *c* (Mylvaganam et al. 1998). In addition to this, we will consider providing links to IM GT/3Dstructure-DB, <http://www.imgt.cines.fr>, which provides standardized annotations for the 3D structures of the IG, TR, MHC, IgSF, and MhcSF, and a query tool for contact analysis (IMGT/StructuralAnalysis) (Kaas et al. 2004).

Immunogenicity and immunodominance of T cell epitopes

In general, the immune responses produced following natural infection or immunization with native antigens does not exploit the full range of epitope possibilities. Rather, only a small fraction of all possible epitopes derived from a given pathogen actually induces an immune response. This phenomenon is known as immunodominance. As stated

above, immunogenicity is not determined by the intrinsic features of an epitope, but rather depends on the context. The following are a number of examples illustrating how context influences immunogenicity.

T cell responses against certain epitopes may go undetected because the epitopes are poorly generated in the breakdown of pathogen-derived proteins within the cells of the host (Chen et al. 2001; Kloetzel 2004; Nakagawa and Rudensky 1999; Stoltze et al. 2000; Yewdell 2001; York et al. 1999). The specificity of the TAP transporter (Daniel et al. 1998; Kloetzel 2004; Lauvau et al. 1999; Uebel and Tampe 1999) and the postproteasomal antigen processing (Reits et al. 2004; Rock et al. 2004) may also play a critical role. The abundance of a particular pathogen-derived protein in intracellular compartments and the presence of proteolytic cleavage sites proximal to, distal to, and within the epitope are just two of many factors that determine this yield (Bergmann et al. 1994; Shastri et al. 1995).

Whether an epitope appears to be immunogenic also depends on the assay method utilized. For example, a given Hepatitis C virus-derived epitope elicited T cells that did not recognize naturally processed Hepatitis C virus antigens expressed by transfected cells, as determined by chromium release killing assays. However, the same CTL line was fully capable of recognizing the same transfected cells if the production of IFN γ was utilized as a readout (Alexander et al. 1998). In a different study (Sette et al. 2001), it was similarly shown that Hepatitis B virus peptide-specific CTL lines apparently incapable of killing endogenous targets, or even producing significant amounts of IFN γ in vitro, were nonetheless capable of controlling viral replication in vivo in Hepatitis B virus transgenic mice.

In other cases, T cells specific for a given epitope are not detectable ("holes in the repertoire"; Klein 1982), because of immunoregulatory phenomena, thymic selection, or peripheral tolerance. In certain cases, the hierarchy of epitope recognition can be altered by prior experiences of the immune system as in memory T cells (Cole et al. 1997). Likewise, the deletion of the immunodominant epitope from a protein antigen can also dramatically change the pattern of epitope recognition, resulting in the activation of T cells against previously subdominant peptides (van der Most et al. 1997). Finally, T cells specific for a given epitope might limit the expansion of other T cell specificities by competition for space or nutrients, competition for cell-cell interactions, or by rapidly eliminating or decreasing the pathogen concentration before other epitope specificities can be effectively stimulated (Chen et al. 2000; Grufman et al. 1999).

The outcome of immunization also varies as a function of immunogen structure (synthetic peptide, recombinant vector, protein, attenuated pathogen) and presence or absence of specific adjuvants. Dose, schedule, and route of immunization also significantly influence the nature of the immune response generated (Brockstedt et al. 1999; Doria-Rose and Haigwood 2003; Horwood and Macfarlane 2002; Imami and Hardy 2003; Murray and Jackson 2002; Scheibenbogen et al. 2003; Sondak and Sosman 2003).

Collectively, these examples illustrate that the failure to detect a response against a given epitope does not necessarily signify that a particular epitope cannot be immunogenic in a different context.

Epitopes recognized in B cell responses

Similar to what is described above for T cell epitopes, not all possible structures expressed by a foreign (or self) protein are recognized as epitopes by B cells and result in antibody responses. Intense investigation has focused on elucidation of the factors that determine antigenicity, immunogenicity, and immunodominance in antibody responses (Aguilar et al. 2001; Atassi et al. 1996; Cleveland et al. 2000; Herkel et al. 2002; Ito et al. 2003; Kanduc et al. 2001; Liang et al. 1999; Musiani et al. 2000; Oleksiewicz et al. 2001; Sobotta et al. 2000; Xiong et al. 2001). It appears that, given appropriate conditions, almost any structure, whether natural or man-made, has the potential to elicit or be recognized by antibody responses.

In general, it appears that B cell epitopes are located on the surface of the structures recognized (protein, virus, bacteria), since antigen processing is not involved and therefore B cells only come in contact with surface components. In addition to structural features, other characteristics and variables play a crucial role in determining immunodominance at the B cell level. For example, the breadth and affinity of antibody responses change dramatically as a function of time and repeated antigenic exposures. In these cases, preferential continued stimulation of high-affinity clones under limiting (decreasing) antigen conditions, combined with the somatic mutation of the immunoglobulin genes encoding the antibody molecule itself (William 1999), leads to affinity maturation. Another mechanism for immunodominance relates to the phenomenon of “original antigenic sin” (Berry et al. 1999; Franks et al. 2003), whereby previous exposure to related (cross-reactive) antigens greatly influences which particular antibody specificities are observed upon subsequent vaccination, infection, or immunization. The presence (or absence) of helper T cells, and their TH1 versus TH2 phenotype, also has a dramatic influence on the type of B cell response observed and on the epitopes recognized (Spellberg and Edwards 2001).

Several additional factors determine recognition of a particular protein within a given pathogen, or of an epitope within a protein. These factors include the type of antigen presenting cell initially encountering the antigen, the presence or absence of “danger” signals, and the presence of opsonizing complement factors and/or antibodies (Aalberse and Platts-Mills 2004; Gallucci and Matzinger 2001; Ishida et al. 2001; Janssens and Beyaert 2003; Jarva et al. 2003; Jiang and Koganty 2003; Kayhty 1998). As in the case of T cell epitopes, these examples illustrate that the immunogenicity or antigenicity of a B cell epitope is highly context-dependent. Therefore, to accurately represent B cell immunogenicity or antigenicity in a database, context-dependent variables must be taken into consideration.

System engineering

The previous sections described the immunological considerations behind our definition of an epitope and the type of immunological information we want to capture. This and the following sections describe how this was translated into a database application design.

We are in the process of developing a formal ontology, which defines a common vocabulary for researchers to share information (Noy and McGuinness 2001). Where possible, we try to build on existing ontologies, such as the IMGT-ONTOLOGY (Giudicelli and Lefranc 1999). This is the first ontology in the domain of immunogenetics and immunoinformatics, which provides a controlled vocabulary and the annotation rules for data, and allows the knowledge management of the IG, TR, MHC, IgSF, and MhcSF of human and other vertebrate species (Lefranc et al. 2005b). The formal specification of the terms to be used in the domain of epitope description, binding, and recognition will be developed to ensure as much as possible accuracy, consistency, and coherence with IMGT (Lefranc et al. 2005a, 2004).

The IEDB classes

The data fields in the IEDB are organized hierarchically into classes and subclasses (Noy and McGuinness 2001). The main classes of the IEDB are defined as “Reference,” “Epitope,” “Binding,” and “Context” (Fig. 2). The class “Reference” contains three main subclasses, each representing the possible sources of data, namely literature, patent, and direct submissions. Literature references will be described by a series of slots (or fields), including the PubMed ID, journal title, and authors. Similarly, in the case of patent references, the patent application number, inventors, title of the application, and related fields are utilized to capture relevant information. Finally, in the case of direct submissions, different fields are utilized to capture the relevant information describing the reference. We anticipate that direct submission of large volumes of data through the NIH-NIAID Large Scale Antibody and T Cell Epitope Discovery Program will be one of the main sources of information for the database.

The next concept of the high level diagram is the “Epitope” class. It is further subdivided in two categories, namely structure and source. Within the Epitope structure category, different fields are meant to capture different types of available information on epitope structure. For example, an epitope could be described by its primary amino acid sequence, by its chemically defined 2D structure, which can be captured in the Simplified Molecular Input Line Entry System (SMILES) (Weininger 1988) notation, or by its 3D structure, which would usually be defined as a subset of atoms in a PDB structure. In the Epitope source category, fields describing the source of the epitope structure are grouped. Examples of these fields are the organism from which the epitope is derived and the

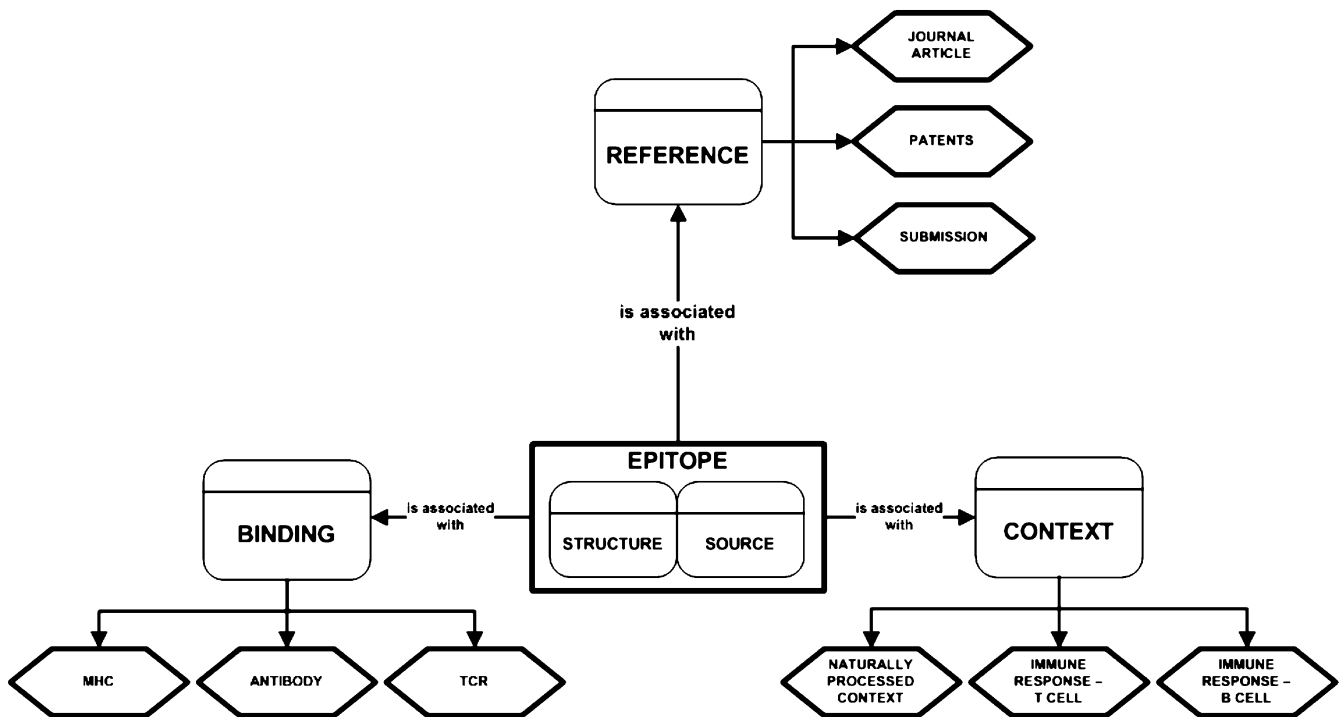


Fig. 2 Class diagram for the Immune Epitope Database and Analysis Resource

relevant GenBank or Swiss-Prot accession numbers identifying the source protein of an epitope.

The “Binding” class captures intrinsic, context-independent information relating to how the structure specified in the “Epitope” class interacts with well-defined receptors of the immune system. Three different subclasses exist: MHC, antibody, and TR. Each subclass encompasses fields relating to the nature of the specific receptors (e.g., PDB, IMGT/3Dstructure-DB, IMGT/GENE-DB, GenBank, Swiss-Prot, or IMGT/LIGM-DB accession numbers), as well as fields relating to the magnitude and modality of interaction such as binding affinity.

The “Context” class is a key novel aspect of our database design. It is organized into three subclasses, including T cell immune responses, naturally processed peptides and B cell immune responses. Within the T cell responses subclass, information fields relating to the TR (of the effector T cells), the immunogen, the administration, and the formulation are found. Other components of this section include fields specifying the source organism from which the MHC of the antigen presenting cells is derived, the antigen, and the type of assay utilized. In general, the subclass “Naturally processed peptides” is similar to the T cell responses category, except for the fact that no effector T cells (and related fields) are present. Likewise, the B cell response subclass is similar to the T cell response subclass, with the main difference that no MHC or antigen-presenting cell is present. This design is robust and flexible and has thus far allowed representation of most immunological situations modeled in our preliminary studies.

As an example of how epitope information would be captured, consider the first sequenced peptide eluted from influenza infected cells (Rotzschke et al. 1990): the refer-

ence fields for this epitope would contain the authors, publication year, journal, and so on. The epitope fields would contain the epitope structure, which is its linear sequence TYQRTRALV, and the epitope source, which is the NP protein of the influenza virus strain A/PR/8/34. In this reference, the epitope is associated with two separate context entries. The “naturally processed” context specifies that this epitope was eluted from mouse cells bearing the MHC allele H2-K^d, and that the antigen containing the epitope in the assay was the influenza virus itself. The “Immune response T cell” context specifies that a CTL response was measured in a ⁵¹Cr release assay, in which the target cells presented the MHC allele H2-K^d, the antigen used was the epitope itself, and the effector cells were a CTL line derived from mice infected with influenza virus.

Guidelines for populating the database

In recent years, different groups have followed different approaches to the discovery of immune epitopes, and have collected privileged data generated by certain assay types (such as direct MHC binding assays, ex vivo T cell detection assays or elution of naturally processed peptides) for the purpose of epitope definition or validation. In a programmatic sense, we believe that selecting data that fit one particular epitope definition or experimental bias is not our prerogative and would be unwise. Rather, we have opted, as described in the preceding sections, for a comprehensive all-inclusive definition. It is our plan to include as many different data sources as possible, and to minimize or exclude bias at the level of database population. A corollary of this approach is that it is necessary to develop

robust and flexible user interfaces, so that the user will not be overwhelmed and will be able to navigate with ease through a large volume of diverse information.

Criteria for data inclusion is publication in peer-reviewed journals, published patents or patent applications, and direct submissions from institutions or companies. Particular attention and priority is given to the inclusion of antibody and T cell epitopes associated with NIAID category A–C pathogens (http://www2.niaid.nih.gov/Bio-defense/bandc_priority.htm), which are considered potential bioterrorism agents. They are divided into categories A–C, where category A is considered the worst potential bioterror threat. NIAID has recently solicited proposals for high-throughput identification of epitopes derived from class A–C pathogens, with 14 distinct contracts having been awarded to different research groups worldwide. These newly established epitope discovery contracts are expected to be a major source of new information, but the database will also include epitopes derived from other emerging/reemerging infectious pathogens, allergens, alloantigens, autoantigens, and model antigens.

All individuals who contribute data or analysis resources to the database will be cited, either by authorship or acknowledgment of their contributions. All of the information contained within the Immune Epitope Database and Analysis Resource, as well as all of the contractor-generated materials (e.g., documentation, software source code, analysis tools, and algorithms) will be made freely available to the entire research community for use, improvement, and publication purposes.

The curation process

We have begun the database population task before deployment of the database because curation of an estimated 80,000 possibly relevant journal articles and 8,000 patent records will be time-consuming. We will define metrics so that the first few months of curation will allow us to refine our labor estimates, and to quantify the value of our proposed curation aids and work flow software support system. Scientific literature will be accessed using PubMed to select articles with a deliberately broad query, then potentially further filtered using a machine-learning model to decrease the number of irrelevant items while maintaining an acceptably low rate of missed relevant items.

The curation workflow on the sorted and queued items is carried out and supervised in a three-tier staffing structure (Fig. 3). The first tier consists of curators who are responsible for the bulk of the data entry and annotation. This data entry team is supervised and coordinated by a second tier of lead curators who oversee data entry and conduct spot checks on the entries to ensure accuracy, consistency, and quality of the annotation. They answer questions that the data entry staff might have and resolve issues related to data curation and annotation. The lead curators also oversee bulk data submissions which will result from either agreement with other existing databases hosting epitope information or data transfer agreements. The third tier,

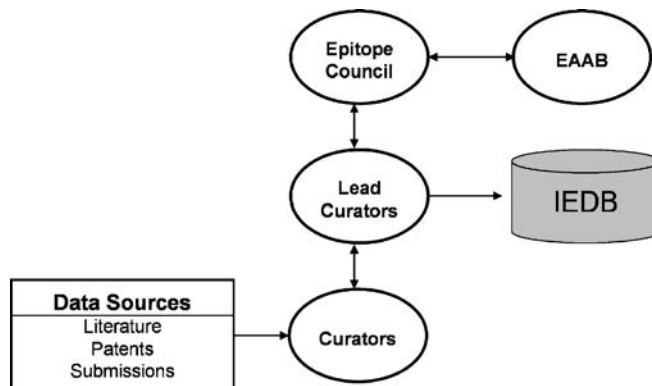


Fig. 3 The curation process

called the Epitope Council, is composed of specialists and recognized authorities in various types of epitopes, who provide input and clarification to the two lower tiers. Approximately ten faculty members of the La Jolla Institute for Allergy and Immunology (LIAI), The Scripps Research Institute (TSRI), and the University of California, San Diego (UCSD) now serve in this capacity on the committee. If the epitope council requires advisory input, it can consult the Epitope Annotation Advisory Board (EAAB) composed of experts in various specific disease systems. The Council may seek the input of specific advisors for any disease-specific issue that might arise that requires specialized know-how and expertise, in-depth knowledge of relevant literature, or an in-depth understanding of issues related to the specific pathology, immunology, and biology of a given pathogen or model organism.

As populating the database begins, the available data are being compiled from various sources (scientific literature, patent applications, and data submissions), starting from the present date and moving backwards in time, with priority given to items containing epitopes from NIAID's category A–C priority pathogen list. Our current rough estimate is that a curator will be able to process several filtered journal articles or patents per day. We have staffed appropriately to accommodate the entire literature and patent backlog within 4 years (approximately 1,000 working days), encompassing new entries, and assuming that a reasonable fraction of new publications will have their data submitted by the authors using a prescribed template format, once the database is on-line. These estimates will be refined as the curation effort proceeds and better metrics become available.

Scientific approach for the development of the analysis resource

Our proposal includes the establishment and maintenance of an Analysis Resource for the Immune Epitope Database. This resource will provide direct online access to various analytical tools where allowed by the tool authors, or links to other public web sites if only links are permitted or feasible. The IEDB team will also develop a pallet of tools

for public access and use. As the goal of this resource is to service the entire scientific community, it is important that the tools provided cover a wide range of research areas relating to epitope discovery and analysis, and that no particular scientific approach has exclusivity. To identify tool candidates, a list of existing tools of interest was initially generated through extensive literature searches and expert input. This list will be periodically revised and expanded, taking advantage of feedback from the scientific community and NIAID. Where desired functionality is not provided by existing tools, the IEDB team will facilitate the development of new tools. An overview of tool categories is provided in Table 2, and a more detailed description of each tool category is given below.

The list of candidate existing tools comprises prediction tools for identifying novel antibody and T cell epitopes from genome and protein sequences. At the level of antibody epitope predictions, standard methods predicting which regions in a protein are likely to be on the surface will be provided, such as hydrophilicity, solvent accessibility, or thermal mobility analysis. Also, tools using various methods for prediction of MHC binding will be provided that include binary main anchor motif scans (both canonical and extended), linear coefficient matrices, neural networks, and other machine learning prediction methods, for as many different MHC as practical at any given point in time. As mentioned above, tools predicting proteasomal processing and TAP transport of T cell epitopes will also be made available. We anticipate that the prediction quality of all of these tools will increase as more data is deposited in the database, which in turn will allow tool developers to train more accurate models.

Apart from these prediction tools, we will provide analytical tool resources to assist in vaccine discovery and development. The population coverage tool will take a set of epitopes with known MHC restriction as an input, and from that project what fraction of a given ethnic population is likely capable to respond to at least one of the epitopes, using MHC allele frequency tables. The conservancy analysis tool will take an epitope and a set of protein sequences as input, and assess the degree of conservancy of the epitope in the sequences, which could be various isolates of the same pathogen or related pathogens. Finally, tools to

visualize data will be provided, such as tools displaying antibody antigen interactions where 3D structural information is either available or can be generated by predictive algorithms.

In terms of how many tools should be hosted, a balance has to be achieved between being too selective, which may leave user demands unaddressed, and hosting too many tools, so that the collection becomes overly redundant and unmanageable. On one hand, resources are required to thoroughly evaluate, host, and maintain the tools, making prioritization necessary in terms of tools actually hosted by the IEDB. On the other hand, links to other websites hosting specific tools can be almost all-inclusive.

To facilitate an objective and transparent choice of which predictive tools should be hosted, the predictions of all candidate tools will be evaluated. This evaluation will be periodically revised as new sets of testing data become available. Most importantly, we plan to make all evaluations (and the methods utilized) publicly available through the IEDB website, and we will encourage all different scientific groups to participate by submitting tools and evaluating data. Such prediction “contests” have had a tremendous positive impact in the evaluation and prediction of protein structure (Fischer and Rychlewski 2003; Moulton et al. 2003; Vajda and Camacho 2004; Venclovas et al. 2003; Wodak and Mendez 2004). The set of evaluations proposed herein represent, to the best of our knowledge, the first attempt at a rigorous and comprehensive evaluation of immune response associated prediction tools. Input, feedback, and collaboration with the scientific community will be a key component of this endeavor.

Discussion

In this work we have described the blueprint for the development of an immune epitope knowledge center, composed of an epitope database and an associated analysis resource. Our vision is based on the definition of epitopes as structures interacting with B or T cell antigen specific receptors, or MHC. In addition to the structures defining the epitopes, we are attempting to also represent context-dependent features of immune epitopes, such as the type of responses evaluated in a particular immunological context.

In terms of developing an Analysis Resource, our approach is inclusive and rigorous. We plan to include a variety of different tools aimed at assisting vaccine design, evaluating immune responses, and identifying T cell and B cell epitopes. Our plans also call for a rigorous, unbiased evaluation of different tools, and open dissemination of the results of the analysis.

In conclusion, our vision is to develop an Immune Epitope Database and Analysis Resource. Our goal is to empower researchers working on immunological evaluations by quickly accessing relevant information and bioinformatics tools. As a result, we hope to facilitate basic research studies and the development and evaluation of prophylactic and therapeutic interventions against new and old, emerging and reemerging disease threats.

Table 2 Tool categories for the analysis resource

Tool category	Examples
T cell epitope identification	Sequence pattern analysis of known epitopes MHC binding predictions Processing predictions (TAP transport, proteasomal cleavage)
B cell epitope identification	Sequence pattern analysis of known epitopes Protein surface predictions
Epitope analysis	Population coverage Conservancy analysis
Visualization	3D viewer of epitope/MHC or antibody/antigen complexes

Acknowledgements This work was supported by the National Institutes of Health Contract HHSN26620040006C.

References

- Aalberse RC, Platts-Mills TA (2004) How do we avoid developing allergy: modifications of the TH2 response from a B-cell perspective. *J Allergy Clin Immunol* 113:983–986
- Aguilar A, Carrazana Y, Duarte CA (2001) Impact of epitope permutations in the antibody response of mice to a multi-epitope polypeptide of the V3 loop of human immunodeficiency virus type 1. *Biomol Eng* 18:117–124
- Alexander J, Del Guercio MF, Fikes JD, Chesnut RW, Chisari FV, Chang KM, Appella E, Sette A (1998) Recognition of a novel naturally processed, A2 restricted, HCV-NS4 epitope triggers IFN-gamma release in absence of detectable cytopathicity. *Hum Immunol* 59:776–782
- Atassi MZ, Dolimbek BZ, Hayakari M, Middlebrook JL, Whitney B, Oshima M (1996) Mapping of the antibody-binding regions on botulinum neurotoxin H-chain domain 855–1296 with antitoxin antibodies from three host species. *J Protein Chem* 15:691–700
- Barber LD, Parham P (1993) Peptide binding to major histocompatibility complex molecules. *Annu Rev Cell Biol* 9:163–206
- Bergmann CC, Tong L, Cua R, Sensintaffar J, Stohlman S (1994) Differential effects of flanking residues on presentation of epitopes from chimeric peptides. *J Virol* 68:5306–5310
- Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, Westbrook J (2000a) The protein data bank and the challenge of structural genomics. *Nat Struct Biol* 7(Suppl):957–959
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000b) The protein data bank. *Nucleic Acids Res* 28:235–242
- Berry JD, Peeling RW, Brunham RC (1999) Analysis of the original antigenic sin antibody response to the major outer membrane protein of *Chlamydia trachomatis*. *J Infect Dis* 179:180–186
- Brockstedt DG, Podsakoff GM, Fong L, Kurtzman G, Mueller-Ruchholtz W, Engleman EG (1999) Induction of immunity to antigens expressed by recombinant adeno-associated virus depends on the route of administration. *Clin Immunol* 92:67–75
- Buslepp J, Wang H, Biddison WE, Appella E, Collins EJ (2003) A correlation between TCR Valpha docking on MHC and CD8 dependence: implications for T cell selection. *Immunity* 19:595–606
- Buus S, Sette A, Colon SM, Miles C, Grey HM (1987) The relation between major histocompatibility complex (MHC) restriction and the capacity of Ia to bind immunogenic peptides. *Science* 235:1353–1358
- Chen W, Anton LC, Bennink JR, Yewdell JW (2000) Dissecting the multifactorial causes of immunodominance in class I-restricted T cell responses to viruses. *Immunity* 12:83–93
- Chen W, Norbury CC, Cho Y, Yewdell JW, Bennink JR (2001) Immunoproteasomes shape immunodominance hierarchies of antiviral CD8(+) T cells at the levels of T cell repertoire and presentation of viral antigens. *J Exp Med* 193:1319–1326
- Cleveland SM, Buratti E, Jones TD, North P, Baralle F, McLain L, McInerney T, Durrani Z, Dimmock NJ (2000) Immunogenic and antigenic dominance of a nonneutralizing epitope over a highly conserved neutralizing epitope in the gp41 envelope glycoprotein of human immunodeficiency virus type 1: its deletion leads to a strong neutralizing response. *Virology* 266:66–78
- Cole GA, Hogg TL, Coppola MA, Woodland DL (1997) Efficient priming of CD8+ memory T cells specific for a subdominant epitope following sendai virus infection. *J Immunol* 158:4301–4309
- Daniel S, Brusica V, Caillat-Zucman S, Petrovsky N, Harrison L, Riganelli D, Sinigaglia F, Gallazzi F, Hammer J, van Endert PM (1998) Relationship between peptide selectivities of human transporters associated with antigen processing and HLA class I molecules. *J Immunol* 161:617–624
- Doria-Rose NA, Haigwood NL (2003) DNA vaccine strategies: candidates for immune modulation and immunization regimens. *Methods* 31:207–216
- Engelhard VH (1994) Structure of peptides associated with class I and class II MHC molecules. *Annu Rev Immunol* 12:181–207
- Engelhard VH, Brickner AG, Zarling AL (2002) Insights into antigen processing gained by direct analysis of the naturally processed class I MHC associated peptide repertoire. *Mol Immunol* 39:127–137
- Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 351:290–296
- Fischer D, Rychlewski L (2003) The 2002 Olympic Games of protein structure prediction. *Protein Eng* 16:157–160
- Fleckenstein B, Jung G, Wiesmuller KH (1999) Quantitative analysis of peptide–MHC class II interaction. *Semin Immunol* 11:405–416
- Flower DR, McSparron H, Blythe MJ, Zygouri C, Taylor D, Guan P, Wan S, Coveney PV, Walshe V, Borrow P, Doytchinova IA (2003) Computational vaccinology: quantitative approaches. *Novartis Found Symp* 254:102–120; discussion 120–5, 216–22, 250–2
- Franks S, Baton L, Tetteh K, Tongren E, Dewin D, Akanmori BD, Koram KA, Ranford-Cartwright L, Riley EM (2003) Genetic diversity and antigenic polymorphism in *Plasmodium falciparum*: extensive serological cross-reactivity between allelic variants of merozoite surface protein 2. *Infect Immun* 71:3485–3495
- Gallucci S, Matzinger P (2001) Danger signals: SOS to the immune system. *Curr Opin Immunol* 13:114–119
- Giudicelli V, Lefranc MP (1999) Ontology for immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics* 15:1047–1054
- Gromme M, Neeffes J (2002) Antigen degradation or presentation by MHC class I molecules via classical and non-classical pathways. *Mol Immunol* 39:181–202
- Grufman P, Wolpert EZ, Sandberg JK, Karre K (1999) T cell competition for the antigen-presenting cell as a model for immunodominance in the cytotoxic T lymphocyte response against minor histocompatibility antigens. *Eur J Immunol* 29:2197–2204
- Hammer J, Valsasnini P, Tolba K, Bolin D, Higelin J, Takacs B, Sinigaglia F (1993) Promiscuous and allele-specific anchors in HLA-DR-binding peptides. *Cell* 74:197–203
- Hennecke J, Carfi A, Wiley DC (2000) Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. *Embo J* 19:5611–5624
- Herkel J, Heidrich B, Nieraad N, Wies I, Rother M, Lohse AW (2002) Fine specificity of autoantibodies to soluble liver antigen and liver/pancreas. *Hepatology* 35:403–408
- Hickman HD, Luis AD, Buchli R, Few SR, Sathiamurthy M, VanGundy RS, Giberson CF, Hildebrand WH (2004) Toward a definition of self: proteomic evaluation of the class I peptide repertoire. *J Immunol* 172:2944–2952
- Horwood F, Macfarlane J (2002) Pneumococcal and influenza vaccination: current situation and future prospects. *Thorax* 57 (Suppl 2):II24–II30
- Imami N, Hardy G (2003) Timing of antiretroviral therapy: an immunological perspective. *J HIV Ther* 8:15–18
- Ishida T, Harashima H, Kiwada H (2001) Interactions of liposomes with cells in vitro and in vivo: opsonins and receptors. *Curr Drug Metab* 2:397–409
- Ito HO, Nakashima T, So T, Hirata M, Inoue M (2003) Immunodominance of conformation-dependent B-cell epitopes of protein antigens. *Biochem Biophys Res Commun* 308:770–776
- Janssens S, Beyaert R (2003) Role of toll-like receptors in pathogen recognition. *Clin Microbiol Rev* 16:637–646
- Jarva H, Jokiranta TS, Wurzner R, Meri S (2003) Complement resistance mechanisms of streptococci. *Mol Immunol* 40:95–107
- Jiang ZH, Koganty RR (2003) Synthetic vaccines: the role of adjuvants in immune targeting. *Curr Med Chem* 10:1423–1439

- Kaas Q, Ruiz M, Lefranc MP (2004) IMGT/3D structure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. *Nucleic Acids Res* 32:D208–D210
- Kanduc D, Lucchese A, Mittelman A (2001) Individuation of monoclonal anti-HPV16 E7 antibody linear peptide epitope by computational biology. *Peptides* 22:1981–1985
- Kayhty H (1998) Immunogenicity assays and surrogate markers to predict vaccine efficacy. *Dev Biol Stand* 95:175–180
- Klein J (1982) *Immunology*. Wiley, New York
- Kloetzel PM (2004) Generation of major histocompatibility complex class I antigens: functional interplay between proteasomes and TPII. *Nat Immunol* 5:661–669
- Latek RR, Unanue ER (1999) Mechanisms and consequences of peptide selection by the I-Ak class II molecule. *Immunol Rev* 172:209–228
- Laemoller SL, Kesmir C, Corbet SL, Fomsgaard A, Holm A, Claesson MH, Brunak S, Buus S (2000) Identifying cytotoxic T cell epitopes from genomic and proteomic information: “the human MHC project.” *Rev Immunogenet* 2:477–491
- Lauvau G, Kakimi K, Niedermann G, Ostankovitch M, Yotnda P, Firat H, Chisari FV, van Endert PM (1999) Human transporters associated with antigen processing (TAPs) select epitope precursor peptides for processing in the endoplasmic reticulum and presentation to T cells. *J Exp Med* 190:1227–1240
- Lefranc M-P, Giudicelli V, Ginestoux C, Bosc N, Folch G, Guiraudou D, Jabado-Michaloud J, Magris S, Scaviner D, Thouvenin V, Combres K, Girod D, Jeanjean S, Protat C, Yousfi Monod M, Duprat E, Kaas Q, Pommier C, Chaume D, Lefranc G (2004) IMGT-ONTOLOGY for immunogenetics and immunoinformatics. *In Silico Biol* 4:17–29
- Lefranc M-P, Clément O, Kaas Q, Duprat E, Chastellan P, Coelho I, Combres K, Ginestoux C, Giudicelli V, Chaume D, Lefranc G (2005a) IMGT-choreography for immunogenetics and immunoinformatics. *In Silico Biol* 5:0006
- Lefranc M-P, Giudicelli V, Kaas Q, Duprat E, Jabado-Michaloud J, Scaviner D, Ginestoux C, Clément O, Chaume D, Lefranc G (2005b) IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* 33(Database Issue):D593–D597
- Lehner PJ (2003) The calculus of immunity: quantitating antigen processing. *Immunology* 18:315–317
- Liang FT, Alvarez AL, Gu Y, Nowling JM, Ramamoorthy R, Philipp MT (1999) An immunodominant conserved region within the variable domain of VlsE, the variable surface antigen of *Borrelia burgdorferi*. *J Immunol* 163:5566–5573
- Madden DR (1995) The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol* 13:587–622
- Margalit H, Altuvia Y (2003) Insights from MHC-bound peptides. *Novartis Found Symp* 254:77–90; discussion 91–101, 216–22, 250–2
- Moult J, Fidelis K, Zemla A, Hubbard T (2003) Critical assessment of methods of protein structure prediction (CASP)-round V. *Proteins* 53(Suppl 6):334–339
- Murray D, Jackson C (2002) A conjugate vaccine for the prevention of pediatric pneumococcal disease. *Mil Med* 167:671–677
- Musiani M, Manaresi E, Gallinella G, Venturoli S, Zuffi E, Zerbini M (2000) Immunoreactivity against linear epitopes of parvovirus B19 structural proteins. Immunodominance of the amino-terminal half of the unique region of VP1. *J Med Virol* 60:347–352
- Mylvaganam SE, Paterson Y, Getzoff ED (1998) Structural basis for the binding of an anti-cytochrome *c* antibody to its antigen: crystal structures of FabE8-cytochrome *c* complex to 1.8 Å resolution and FabE8 to 2.26 Å resolution. *J Mol Biol* 281:301–322
- Nakagawa TY, Rudensky AY (1999) The role of lysosomal proteinases in MHC class II-mediated antigen processing and presentation. *Immunol Rev* 172:121–129
- Noy NF, McGuinness DL (2001) Ontology development 101: a guide to creating your first ontology. Stanford Knowledge Systems Laboratory technical report KSL-01-05 and Stanford Medical Informatics technical report
- Oleksiewicz MB, Botner A, Toft P, Normann P, Storgaard T (2001) Epitope mapping porcine reproductive and respiratory syndrome virus by phage display: the nsp2 fragment of the replicase polyprotein contains a cluster of B-cell epitopes. *J Virol* 75:3277–3290
- Rammensee HG (1995) Chemistry of peptides associated with MHC class I and class II molecules. *Curr Opin Immunol* 7:85–96
- Rammensee HG, Falk K, Rotzschke O (1993) Peptides naturally presented by MHC class I molecules. *Annu Rev Immunol* 11:213–244
- Reits E, Neijssen J, Herberths C, Benckhuijsen W, Janssen L, Drijfhout JW, Neeffjes J (2004) A major role for TPII in trimming proteasomal degradation products for MHC class I antigen presentation. *Immunology* 20:495–506
- Rock KL, York IA, Goldberg AL (2004) Post-proteasomal antigen processing for major histocompatibility complex class I presentation. *Nat Immunol* 5:670–677
- Rotzschke O, Falk K, Deres K, Schild H, Norda M, Metzger J, Jung G, Rammensee HG (1990) Isolation and analysis of naturally processed viral peptides as recognized by cytotoxic T cells. *Nature* 348:252–254
- Rotzschke O, Falk K, Stevanovic S, Jung G, Walden P, Rammensee HG (1991) Exact prediction of a natural T cell epitope. *Eur J Immunol* 21:2891–2894
- Saveanu L, Fruci D, van Endert P (2002) Beyond the proteasome: trimming, degradation and generation of MHC class I ligands by auxiliary proteases. *Mol Immunol* 39:203–215
- Scheibenbogen C, Schadendorf D, Bechrakis NE, Nagorsen D, Hofmann U, Servetopoulou F, Letsch A, Philipp A, Foerster MH, Schmittel A, Thiel E, Keilholz U (2003) Effects of granulocyte-macrophage colony-stimulating factor and foreign helper protein as immunologic adjuvants on the T-cell response to vaccination with tyrosinase peptides. *Int J Cancer* 104:188–194
- Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* 50:201–212
- Sette AD, Oseroff C, Sidney J, Alexander J, Chesnut RW, Kakimi K, Guidotti LG, Chisari FV (2001) Overcoming T cell tolerance to the hepatitis B virus surface antigen in hepatitis B virus-transgenic mice. *J Immunol* 166:1389–1397
- Shastri N, Serwold T, Gonzalez F (1995) Presentation of endogenous peptide/MHC class I complexes is profoundly influenced by specific C-terminal flanking residues. *J Immunol* 155:4339–4346
- Shastri N, Serwold T, Paz P (1998) Reading within the lines: naturally processed peptides displayed by MHC class I molecules. *Curr Opin Immunol* 10:137–144
- Sobotta D, Sominskaya I, Jansons J, Meisel H, Schmitt S, Heermann KH, Kaluza G, Pumpens P, Gerlich WH (2000) Mapping of immunodominant B-cell epitopes and the human serum albumin-binding site in natural hepatitis B virus surface antigen of defined genosubtype. *J Gen Virol* 81:369–378
- Sondak VK, Sosman JA (2003) Results of clinical trials with an allogenic melanoma tumor cell lysate vaccine: melacine. *Semin Cancer Biol* 13:409–415
- Spellberg B, Edwards JE Jr (2001) Type 1/Type 2 immunity in infectious diseases. *Clin Infect Dis* 32:76–102
- Stevanovic S (2002) Structural basis of immunogenicity. *Transpl Immunol* 10:133–136
- Stevanovic S, Lemmel C, Hantschel M, Eberle U (2003) Generating data for databases—the peptide repertoire of HLA molecules. *Novartis Found Symp* 254:143–155; discussion 155–64, 216–22, 250–2
- Stoltze L, Nussbaum AK, Sijts A, Emmerich NP, Kloetzel PM, Schild H (2000) The function of the proteasome system in MHC class I antigen processing. *Immunol Today* 21:317–319
- Uebel S, Tampe R (1999) Specificity of the proteasome and the TAP transporter. *Curr Opin Immunol* 11:203–208
- Vajda S, Camacho CJ (2004) Protein-protein docking: is the glass half-full or half-empty? *Trends Biotechnol* 22:110–116

- Van Bleek GM, Nathenson SG (1990) Isolation of an endogenously processed immunodominant viral peptide from the class I H-2Kb molecule. *Nature* 348:213–216
- van der Merwe PA, Davis SJ (2003) Molecular interactions mediating T cell antigen recognition. *Annu Rev Immunol* 21:659–684
- van der Most RG, Concepcion RJ, Oseroff C, Alexander J, Southwood S, Sidney J, Chesnut RW, Ahmed R, Sette A (1997) Uncovering subdominant cytotoxic T-lymphocyte responses in lymphocytic choriomeningitis virus-infected BALB/c mice. *J Virol* 71:5110–5114
- Van Kaer L (2002) Major histocompatibility complex class I-restricted antigen processing and presentation. *Tissue Antigens* 60:1–9
- Van Regenmortel MHV (1996) Mapping epitope structure and activity: from one-dimensional prediction to four-dimensional description of antigenic specificity. *Methods* 9:465–472
- Venclovas C, Zemla A, Fidelis K, Moulton J (2003) Assessment of progress over the CASP experiments. *Proteins* 53(Suppl 6):585–595
- Weininger D (1988) SMILES 1. Introduction and encoding rules. *J Chem Inf Comput Sci* 28
- William PE (1999) *Fundamental immunology*. Lippincott, Williams and Wilkins, Philadelphia
- Wodak SJ, Mendez R (2004) Prediction of protein–protein interactions: the CAPRI experiment, its evaluation and implications. *Curr Opin Struct Biol* 14:242–249
- Xiong Z, Farilla L, Guo J, McLachlan S, Rapoport B (2001) Does the autoantibody immunodominant region on thyroid peroxidase include amino acid residues 742–771? *Thyroid* 11:227–231
- Yewdell JW (2001) Not such a dismal science: the economics of protein synthesis, folding, degradation and antigen processing. *Trends Cell Biol* 11:294–297
- York IA, Goldberg AL, Mo XY, Rock KL (1999) Proteolysis and class I major histocompatibility complex antigen presentation. *Immunol Rev* 172:49–66