JOURNAL OF MOLECULAR RECOGNITION
J. Mol. Recognit. 2007; **20**: 75–82
Published online 5 January 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI:10.1002/jmr.815

Review

Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools†

Jason A. Greenbaum[1]*, Pernille Haste Andersen[2], Martin Blythe[3], Huynh-Hoa Bui[1], Raul E. Cachau[4], James Crowe[5], Matthew Davies[6], A. S. Kolaskar[7], Ole Lund[2], Sherrie Morrison[8], Brendan Mumey[9], Yanay Ofran[10], Jean-Luc Pellequer[11], Clemencia Pinilla[12], Julia V. Ponomarenko[13], G. P. S. Raghava[14], Marc H. V. van Regenmortel[15], Erwin L. Roggen[16], Alessandro Sette[1], Avner Schlessinger[17], Johannes Sollner[18], Martin Zand[19] and Bjoern Peters[1]

[1]Immune Epitope Database and Analysis Resource (IEDB), La Jolla Institute for Allergy and Immunology, La Jolla, California, USA
[2]Immunological Bioinformatics Group, Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark, Lyngby, Denmark
[3]Bioinformatics Department, The Edward Jenner Institute for Vaccine Research, Compton, Newbury, Berkshire, England, UK
[4]NCI-FCRDC, Frederick Biomedical Supercomputer Center, Frederick, Maryland, USA
[5]Departments of Pediatrics, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[6]Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford, England, UK
[7]Bioinformatics Centre, University of Pune, Maharashtra, India
[8]University of California in LA, Los Angeles, California, USA
[9]Computer Science Department, Montana State University, Bozeman, Montana, USA
[10]Department of Biochemistry, Tel Aviv University, Faculty of Life Sciences, Ramat Aviv, Israel
[11]CEA Valrhô—Centre de Marcoule, DSV/DIEP/SBTN, Bat 170 BP17171, 30207 Bagnols sur Cèze Cedex, France
[12]Torrey Pines Institute for Molecular Studies, 3550 General Atomics Ct, San Diego, California, USA
[13]San Diego Supercomputer Center, University of California, San Diego, La Jolla, California, USA
[14]Bioinformatics Centre, Institute of Microbial Technology, Sector 39A, Chandigarh, India
[15]Ecole Supérieure de Biotechnologie de Strasbourg-Parc d'innovation, Illkirch Cedex 67412, France
[16]Department of Pharma Protein Development, Molecular Biotechnology, Novozymes AS, Smoermosevej 11, 2880 Bagsvaerd, Denmark
[17]Columbia University Center for Computational Biology, Department of Biochemistry & Molecular Biophysics, New York, New York, USA
[18]Emergentec, 1010 Vienna, Austria
[19]Center for Biodefense Immune Modeling, University of Rochester Medical Center, Rochester, New York, USA

A B-cell epitope is the three-dimensional structure within an antigen that can be bound to the variable region of an antibody. The prediction of B-cell epitopes is highly desirable for various immunological applications, but has presented a set of unique challenges to the bioinformatics and immunology communities. Improving the accuracy of B-cell epitope prediction methods depends on a community consensus on the data and metrics utilized to develop and evaluate such tools. A workshop, sponsored by the National Institute of Allergy and Infectious Disease (NIAID), was recently held in Washington, DC to discuss the current state of the B-cell epitope prediction field. Many of the currently available tools were surveyed and a set of recommendations was devised to facilitate improvements in the currently existing tools and to expedite future tool development. An underlying theme of the recommendations put forth by the panel is increased collaboration among research groups. By developing common datasets, standardized data formats, and the means with which to consolidate information, we hope to greatly enhance the development of B-cell epitope prediction tools. Copyright © 2007 John Wiley & Sons, Ltd.

Keywords: **B-cell epitopes; epitope prediction; algorithms; immunology databases; data standardization; bioinformatics; software tools; tool development**

Received 21 November 2006; accepted 28 November 2006

INTRODUCTION

Antibodies are proteins produced by B-cells in response to immunogenic substances such as viruses, allergens, and vaccines. B-cell, or antibody, epitopes are the molecular structures within an antigen that make specific contacts with residues of soluble and membrane-bound antibody molecules (See Figure 1 for a schematic introduction to key terms). For various immunological applications, a computational prediction of epitopes in an antigen is highly desirable. Many such tools exist that mainly base their predictions on shared amino acid characteristics of known B-cell epitopes in protein antigens. However, in contrast to predictions used to identify T-cell epitopes (Brusic *et al.*,

*Correspondence to: J. A. Greenbaum, Immune Epitope Database and Analysis Resource (IEDB), La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA. E-mail: jgbaum@liai.org
†The manuscript was compiled based on discussions and contributions of all authors at the workshop. All authors contributed edits and approved the final manuscript.

Copyright © 2007 John Wiley & Sons, Ltd.

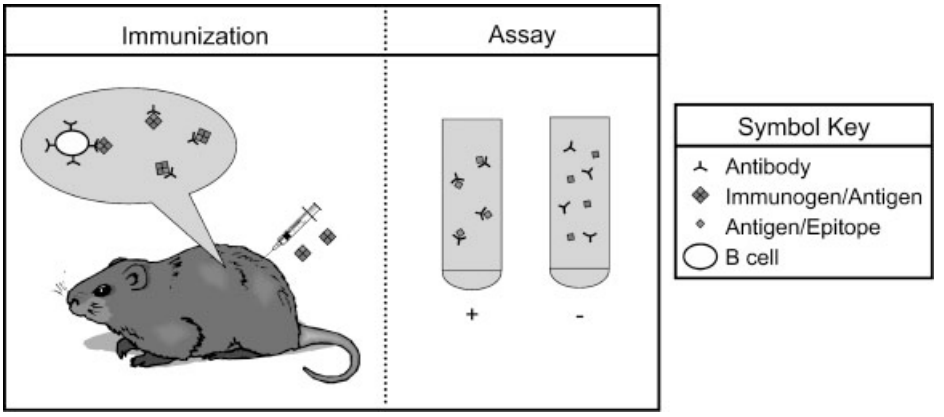

Figure 1. B-cell epitope mapping schematic. The figure schematically depicts an epitope mapping experiment. In an immunization (left panel), the host species is exposed to a substance that elicits the production of antibodies by B-cells. This substance is referred to as the immunogen, while the substance recognized by antibodies is referred to as the antigen. While antigen and immunogen are typically the same substance during the course of an infection, they are often distinct in an epitope mapping experiment. Such an experiment is depicted in the right panel, where different fragments or mutants of the immunogen are used as antigens in order to identify the specific antibody binding sites, which are referred to as epitopes.

2004; De Groot, 2006), the quality of these predictions is widely considered to be too poor to be employed as a reliable tool by immunologists (Blythe and Flower, 2005). From this point forward, the term 'epitope' shall refer to a B-cell epitope, unless otherwise specified.

In a workshop sponsored by the National Institute for Allergy and Infectious Diseases (NIAID) on 7 and 8 September 2006, a panel of immunologists and bioinformaticians convened in Washington, DC to discuss the current state and future directions for the epitope prediction field. The specific goals of the workshop were to review currently available epitope prediction tools, agree upon metrics to evaluate tool performance, identify a body of relevant training and test data, and develop recommendations for advertising and implementing these suggestions for the broader research community. This meeting report is meant to publicize a summary of the presentations and discussions at the workshop, and elicit community feedback, via a web-based forum (http://www.immuneepitope.org/jive/), on the specific proposals and recommendations.

B-cell epitope terminology

Dr Marc H. V. van Regenmortel opened the workshop with an overview of epitope terminology issues. It was noted that researchers classify epitopes as either continuous or discontinuous (Figure 2), but many so-called discontinuous epitopes consist of stretches of several consecutive amino acids that could, in some cases, be considered continuous epitopes in their own right (van Regenmortel, 2006). Further, the distinction between linear and conformational epitopes is problematic, since linear peptides necessarily adopt a particular three-dimensional conformation that is recognized by their cognate antibody. Dr van Regenmortel also emphasized the need to clarify the purpose of making a specific epitope prediction, and how this clarification could direct selection of the most appropriate prediction tool or development of a new tool, as needed. For instance, if the purpose is to predict epitopes as vaccine candidates, then the predicted epitopes should elicit antibodies that recognize the vaccine target and could possibly provide protection from infection (van Regenmortel, 2001). However, if the purpose is to predict epitopes that could be used to replace complete infectious antigens in diagnostic immunoassays, the predicted epitopes should be able to react with antibodies found in hosts infected with the pathogenic agent. Since any given purpose may apply only to a subset of epitopes, the main goal of the prediction should be taken into account in selecting an appropriate dataset for tool training and evaluation.

The influence of antibody structure on its function was described in the context of responses to rotavirus infections by Dr James Crowe (Crowe *et al.*, 2001). Central to the notion of an epitope is the understanding that epitopes are context dependent. That is, they are dependent entities and therefore cannot exist without a corresponding antibody. At the same time, genes encoding antibodies undergo a rapid process of somatic hypermutation during germinal center reactions resulting in variations of their binding site. It may therefore be useful to think of a protein surface as a continuous landscape of epitopic regions without well-defined borders. Under a given set of circumstances, any region of this landscape can behave as an epitope. By this definition, a very large ensemble of epitopes exists on any given protein surface. To complicate matters further, a portion of a protein surface that acts as an epitope under one set of circumstances (e.g., assay type, antibody concentration, antibody species specificity) will not necessarily behave as an epitope under another set. Therefore, binary classification of antigen regions into epitopes or non-epitopes may not accurately reflect biological reality.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/jmr

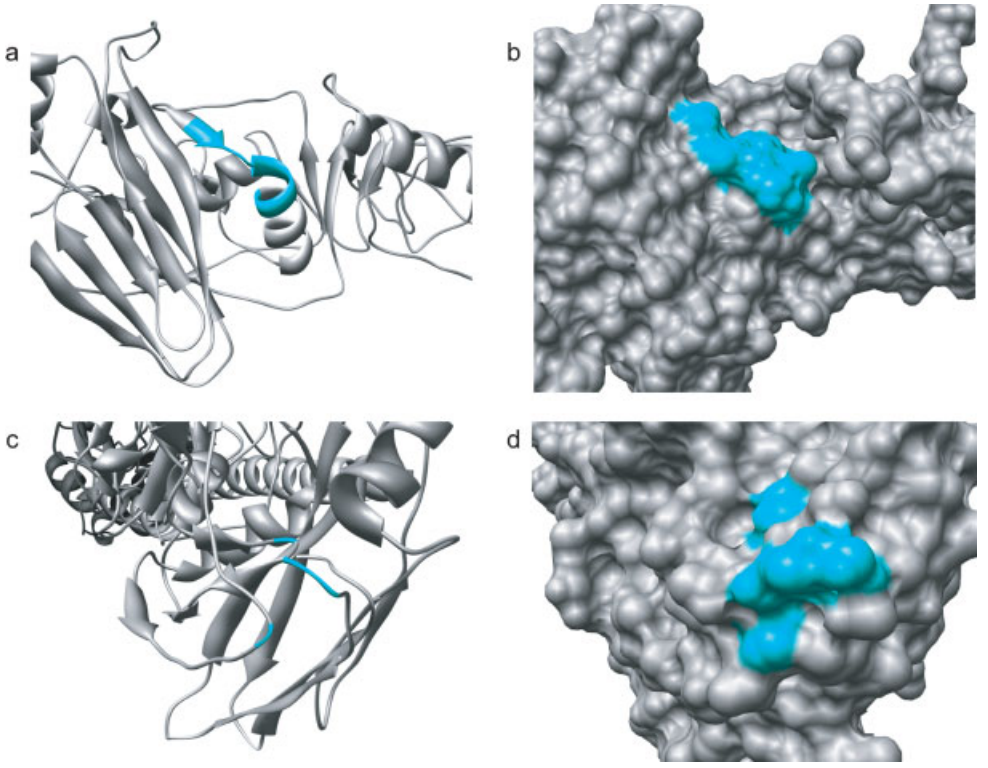

Figure 2. Continuous versus discontinuous epitopes. This figure contrasts the difference between a continuous and discontinuous epitope in Hemaglutinin precursor protein (Swiss-Prot ID: P03435, PDB code: 1HA0) of Influenza A virus. In each rendering, the residues that are part of the epitope are colored in cyan, while the remaining residues are gray. (a) Continuous epitope (Churchill *et al.*, 1994) represented as ribbons. Here, the epitope sequence lies in one region of the protein. This is the minimal sequence required for recognition by the antibody. (b) Surface rendering of epitope shown in (a). (c) Discontinuous epitope (Underwood, 1984) represented as ribbons. In this case, the epitope consists of residues that are distant in the primary sequence, but close when the protein is folded into its native three-dimensional structure. All of the residues are required for recognition by the antibody and thus are not epitopes on their own. (d) Surface rendering of epitope shown in (c). This rendering elucidates how residues that are distant in the primary sequence can be part of the same epitope.

B-cell epitope datasets currently available and under development

Defining the requirements for assembling relevant datasets was a recurring theme at the workshop. Martin Blythe described the data collected in the AntiJen database (Toseland *et al.*, 2005) (Table 1), which consists of 3541 datapoints and is one of the largest available resources of linear and conformational epitope information. A dataset consisting of 50 linear epitope-mapped protein sequences was derived from the database (Blythe and Flower, 2005). Several other groups have utilized these information resources for tool development and testing purposes. Other public data sources that were used for tool development and evaluation include: the comparably small, but high-quality, dataset of Dr Jean-Luc Pellequer (Pellequer *et al.*, 1993), consisting of 82 well-defined continuous epitopes across 14 proteins that have been studied extensively; the HIV Molecular Immunology Database, hosted at Los Alamos National Labs (Korber *et al.*, 2005); and the Bcipep database by Dr G. P. S. Raghava's group (Saha *et al.*, 2005), containing 3031 epitopes.

The approach towards assembling epitope datasets from Immune Epitope Database (IEDB) (Peters *et al.*, 2005) was presented by Dr Alessandro Sette. The IEDB contains highly annotated T- and B-cell epitopes, from which users can select datasets relevant to their particular purpose. For example, epitopes identified by immunization with the native antigen and assayed with peptides can be considered separately from those immunized with peptides and assayed with native antigen. Along with epitopes identified using functional assays, the IEDB includes curated data on epitopes inferred from three-dimensional structures of antigen–antibody complexes available in the Protein Data Bank (PDB) (Berman *et al.*, 2000). The database also includes a large volume of negative data; that is, peptide and protein sequences experimentally shown not to be recognized by antibodies in a particular assay. It should be emphasized that these regions may, in fact, behave as epitopes when sampled under a different set of conditions, or with a different population of antibodies. To provide user-friendly datasets for epitope prediction tool development and testing, the IEDB team is in the process of assembling all relevant, non-redundant data into several customized datasets, which will be available to the public by mid-2007.

B-cell epitope prediction tools

Currently available tools and their evaluations were discussed by Dr Jean-Luc Pellequer. Specifically, the

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/jmr

Table 1. Tools, databases, and datasets

Name	URL/email	Description
ABCpred	http://www.imtech.res.in/raghava/abcpred	Sequence-based machine-learning tool for the prediction of continuous epitopes
AntiJen*	http://www.jenner.ac.uk/AntiJen/	Database of binding data for various types of proteins, including B-cell epitopes and antibodies
Bcipep*	http://www.imtech.res.in/raghava/bcipep/	Database of B-cell epitopes of varying immunogenicity
Bepipred	http://www.cbs.dtu.dk/services/BepiPred	Sequence-based tool for the prediction of continuous epitopes
BEPITOPE	jlpellequer@cea.fr	Sequence-based tool for the prediction of continuous epitopes
CEP	http://bioinfo.ernet.in/cep.htm	Structure-based tool for the prediction of continuous and discontinuous epitopes
DiscoTope	http://www.cbs.dtu.dk/services/DiscoTope	Sequence/structure-based tool for the prediction of discontinuous epitopes
EMT	elro@novozymes.com	Phage-display based tool for the prediction of continuous and discontinuous epitopes
EPIMAP	mumey@cs.montana.edu	Phage-display based tool for the identification of discontinuous epitopes
Epitome	http://www.rostlab.org/services/epitome	Database of antigenic residues and interacting antibodies including detailed descriptions and visualization capabilities
HIV database*	http://hiv-web.lanl.gov/content/immunology/	Database of HIV-specific immune epitopes
IEDB**	http://www.immuneepitope.org	Database of T- and B-cell epitopes and non-epitopes
IEDB B-cell epitope tools	http://www.immuneepitope.org/tools/bcell/iedb_input	Sequence-based tool for the prediction of continuous epitopes
Pellequer dataset	jlpellequer@cea.fr	Dataset consisting of 82 epitopes in 14 protein sequences
SPA	johannes.soellner@emergentec.com	Program for the parametrization of peptide sequences. Intended for use with machine-learning algorithms

This table summarizes the available B-cell epitope datasets, databases, and prediction tools. Where URLs are not available, author's email addresses are supplied.
* A dataset for tool training and testing has been assembled from this database and is available at the listed URL.
** A dataset for tool training and testing is being assembled from this database and will be available by mid-2007.

distinction between a *database* and a *dataset* was elucidated. The former serves as the data warehouse and contains all potentially relevant data, while the latter is a smaller subset of data intended for tool development purposes. Following this general introduction, various currently available tools for epitope prediction were presented and are briefly summarized and loosely categorized here: sequence-based tools—tools that take only amino acid sequence information as input; structure-based tools—tools that require some form of three-dimensional structure information; and tools incorporating data from phage-display experiments.

SEQUENCE-BASED PREDICTION TOOLS

The 'classical' approach to epitope prediction is to utilize amino acid propensity scales such as hydrophilicity or chain flexibility to identify regions in antigens that are likely to contain epitopes. One such approach is performed by the BEPITOPE tool (Odorico and Pellequer, 2003), presented by Dr Pellequer, which makes use of the predictive power of a consensus voting method based on the prediction of turns in proteins. A consensus method that combines several different prediction methods is also available in BEPITOPE. Additionally, several scale-based epitope prediction tools have been made available on the IEDB website. Evaluation of these tools, using datasets from Pellequer (Pellequer *et al.*, 1993) and the AntiJen (Toseland *et al.*, 2005) and HIV databases, gave area under the receiver operating characteristic curve (A_{ROC}) (Swets, 1988) values around 0.60, similar to those reported by other groups for amino acid scales. Briefly, A_{ROC} scores range from 0 to 1, with a score of 0.5 equal to random discrimination, and 1 equal to perfect performance (Lund *et al.*, 2005). The most detailed evaluation of such scales was presented by Martin Blythe

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/jmr

(Blythe and Flower, 2005), who tested 484 amino acid scales from the AAindex database (Kawashima *et al*., 1999). The combination of scales and experimentation with several machine-learning algorithms showed little improvement over single scale-based methods, which were considered to perform inadequately.

Machine-learning tools attempt to extract characteristics of an epitope from a set of learning examples, and generalize them in a classification algorithm. One such approach using an artificial neural network (ANN) was applied by Saha and Raghava (2006) in the ANN-Based B-cell Epitope Prediction (ABCpred) algorithm. The system was trained on 700 epitopes from the Bcipep database and 700 randomly selected peptides represented by amino acid sequences of lengths varying between 10 and 20 amino acids. Employing fivefold cross-validation on this dataset, the method achieved a maximum accuracy of ~66%. Bepipred (Larsen *et al*., 2006), an algorithm that combines scores from the Parker hydrophilicity scale (Parker *et al*., 1986) and a hidden Markov model (HMM) trained on linear epitopes, was presented by Dr Ole Lund. The method shows a small, but significant, increase in A_{ROC} over earlier scale-based methods. The sequence parametrizer (SPA) algorithm (Sollner, 2006; Sollner and Mayer, 2006), along with its associated machine-learning methods, was presented by Dr Johannes Sollner. In addition to using the common single amino acid propensity scales, this method also incorporates neighborhood parameters reflecting the probability that a given stretch of amino acids exists within a predefined proximity of a specific amino acid residue. Training and testing on epitope sequences pulled from a high-quality proprietary database, as well as several publicly accessible databases, yields a degree of accuracy that is greatly increased over single-parameter methods.

STRUCTURE-BASED PREDICTION TOOLS

Several groups presented novel tool development efforts taking advantage of three-dimensional structures of antigen–antibody complexes available in the PDB. The DiscoTope algorithm (Haste Andersen *et al*., 2006), presented by Pernille Haste Andersen, is a sequence-structure hybrid method. A log-odds probability matrix of each amino acid residue taking part in an antigenic interaction was calculated based on data compiled from antigen–antibody complexes. This innovative algorithm combines scores from this matrix with a measure of the surface area to predict the location of discontinuous epitope residues and does so with a fair degree of accuracy. Using the A_{ROC} as an indicator, the algorithm achieves a score of 0.711.

Another method employing the three-dimensional structure of the antigen for prediction of epitopic regions was presented by Dr A. S. Kolaskar. The conformational epitope prediction (CEP) (Kulkarni-Kale *et al*., 2005) server calculates the relative accessible surface area (RSA) for each residue in the structure and determines which regions of the protein molecule are sufficiently exposed to act as antigenic determinants. Additionally, regions distant in the primary sequence, but close in three-dimensional space are condensed into one epitope. To test the tool, in a dataset consisting of 63 antigen–antibody complexes, the algorithm correctly identified 76% of the epitopic residues.

Antigen–antibody complexes from the PDB were annotated and compiled into a database (Epitome) (Schlessinger *et al*., 2006) by Drs Yanay Ofran and Avner Schlessinger. Using structural alignments of antibodies bound to different antigens, the antibody residues that are part of the complementarity-determining region (CDR) could be defined. Contacts to this region are thought to play a role in the specific recognition of an epitope by an antibody. Consolidation of data in this manner allowed the authors to observe several interesting features at the sequence level of antigens. Specifically, compared to standard protein–protein interactions, the antigenic residues involved in antigen–antibody interactions exhibit a lower degree of conservation than residues involved in other protein–protein interactions. This finding, coupled with several other sequence factors, could help to predict antigenic sites through sequence analysis.

Non-redundant datasets of representative three-dimensional structures of protein antigens and antigen–antibody complexes from the PDB were assembled by Dr Julia Ponomarenko. These datasets were used to determine the sensitivity of antibody binding site prediction of several tools developed for various purposes. Established tools for epitope prediction (CEP), protein–protein interaction prediction [PPI-PRED (Bradford and Westhead, 2005), ProMate (Neuvirth *et al*., 2004)], and protein–protein docking [ClusPro (Comeau *et al*., 2004), PatchDock (Duhovny *et al*., 2002)] were included in the study. While in this comparison, varying evaluation metrics showed that none of the tested methods demonstrated a very high degree of accuracy, the protein–protein docking algorithms yielded the highest sensitivity (56%). This underperformance can most likely be attributed to the fact that most of these methods were not designed for the purpose of epitope predictions. Consistent with the findings of the DiscoTope and Epitome studies, the comparison of sequence characteristics within epitopes to other surface residues, revealed a significantly lower degree of conservation, as well as significant differences in the frequency of individual residues and secondary structural elements.

TOOLS INCORPORATING PHAGE-DISPLAY DATA

One important and pervasive experimental technique to map epitopes is the use of phage-display libraries (Smith and Petrenko, 1997). By selecting phages from a library for their ability to bind antibodies specific for a known antigen, linear peptide sequences that cross-react with these antibodies, commonly referred to as mimotopes (Meloen *et al*., 2000), can be discovered. The exact linear sequence identified will generally not exist as a linear sequence in the antigenic protein. The identification of the region of the antigen mimicked by one of these peptides therefore requires special tools that, in turn, provide insight into antigen–antibody binding interactions.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/jmr

Dr Erwin L. Roggen presented work on his Epitope Mapping Tool (EMT) (Batori *et al*., 2006) that scrutinizes an antigen for the presence or absence of amino acid motifs commonly found in epitopes. Epitopic motifs were compiled by the alignment of sequences captured by the competitive immunoscreening of phage-display libraries. Combining the detection of these epitopic motifs with evaluations of the relative surface accessibility of residues in an antigen yielded predictions of the antigenicity of residues in food (e.g., lysozyme, Ara h 1, Ara h 2, Ara h 3), environmental (Der p 1, Bet v 1), and technical (industrial proteases) allergens that were in close agreement with experimental data.

The EPIMAP (Mumey *et al*., 2003, 2006) method was presented by Dr Brendan Mumey. In this method, mimotopes discovered by phage display are individually aligned against their parent antigen via a dynamic programming algorithm. Resultant plots of the frequency of alignment for each residue of the parent protein can reveal segments that are close in three-dimensional space in the native fold. Application of this technique to IL-10 mimetic peptides has yielded residue classifications consistent with an experimentally mapped epitope.

Peptide sequences collected from the biopanning of phage libraries were mapped to three-dimensional structures by Dr Raul E. Cachau (Cachau *et al*., 2003). This was accomplished by applying feedback restrained molecular dynamics (FRMD) to the peptide pool, under the assumption that all the mimotopes should be able to adopt a similar conformation. Validation was performed on an erythropoietin analog and the model was shown to be in good agreement with available structural data.

WORKSHOP RECOMMENDATIONS

Following formal presentations, workshop participants were asked to provide recommendations on the following topics:

- Assembly of ideal datasets, including detailed information regarding dataset characteristics.
- Methods and metrics for tool evaluation.
- Resource needs and action items.

Characteristics of an ideal dataset

On a number of discussed issues, a broad but not necessarily unanimous consensus could be reached. One such issue was the assembly of a high-quality dataset for the training and testing of current and novel prediction algorithms. The quality of the underlying dataset is intrinsically linked to the predictive ability of the tools. In turn, the accessibility to a database of well-curated epitope data is essential to the development of a reliable dataset. This distinction between the database, which includes all possible information, and the dataset, which is a subset of relevant data from the database, is subtle, but critical to comprehend. Ideally, a database should include the following information: sequence information characterizing the epitope and antigen, the relationship between recognition of epitope and antigen (i.e., does immunization with the epitope produce antibodies capable of recognizing the native antigen and, conversely, does immunization with the antigen produce antibodies that can recognize epitopes contained within the antigen); the conformation in which the antigen is tested (i.e., native vs. non-native); the location of the epitope within the three-dimensional structure of the source antigen (if available); the host species producing the antibody, a classification of the antibody as functional (i.e., neutralizing or protective); and quantitative data on the interaction between epitope and antibody (e.g., binding affinity, frequency of recognition, etc.). Importantly, a database should also capture negative data, (i.e., regions of antigens that were experimentally tested and shown *not* to be recognized by antibodies). It should be stressed that incorporation of negative data in the training and testing datasets is as vital as the inclusion of epitope data, and should therefore be adequately represented. This type of database allows for the assembly of several training and testing datasets, one for use with each particular type of application. As alluded to earlier, the purpose of any specific prediction algorithm will dictate the proper dataset upon which it should be trained and evaluated.

As the datasets naturally expand, the establishment of a central public repository is desirable for direct comparison of tool performance, and to avoid unnecessary duplication of work. The IEDB can provide non-exclusive online hosting for such datasets, allowing all tool developers and immunologists to deposit their datasets onto a central server. At the same time, all content from the IEDB is free for download and redistribution in any form. Several databases and datasets, with unique and useful characteristics, were created before the IEDB, and likely will continue to be created and maintained. The IEDB developers propose that the IEDB can act as a central repository to facilitate the exchange of information by providing access to these datasets and proper credit to their creators. This arrangement will allow for several publicly accessible datasets to evolve in parallel, eventually leading to a community consensus on the best and most appropriate datasets for tool training and evaluation.

Metrics for tool evaluation

The purpose of tool evaluations is twofold: (1) Tool developers need a statistically sound assessment of tool prediction quality that allows for the comparison of current tools and aids in the development of new ones. The broad array of metrics currently used to report tool performance illustrates the need for a common measure. It was generally agreed that the most appropriate metric for tool performance evaluation is the A_{ROC}. This measure has the advantage of being non-parametric and does not require a threshold for determining a positive prediction. The range of values is easily interpreted, with a score of 0.5 equal to random performance and 1 corresponding to perfect performance. (2) Tool users need an easily understandable assessment of the level of accuracy that can be expected when making a prediction using a giving tool. This need makes additional metrics necessary, such as sensitivity, specificity, and positive predictive value. In addition to providing metrics that assess tool performance based upon a standard dataset, a

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/jmr

narrative must be presented to tool users that explains their meaning and implications for biological applications.

The automated evaluation of tools on a regular basis using a standard dataset and set of metrics was discussed. This approach was generally thought to be a useful pursuit, as it has been an asset to the protein secondary structure prediction (Koh *et al*., 2003) and transmembrane helix prediction communities (Kernytsky and Rost, 2003). To employ an automated server of this type, it is imperative that each tool incorporates the ability to import and export data in a common format. This standardization would not only facilitate the development of an evaluation server, but also the development of a metaserver for the user community that would act as a front end to the tool set. The ability to run related tools from one central server has proved invaluable in other fields [e.g., protein structure prediction (Bujnicki *et al*., 2001) and protein domain prediction (Saini and Fischer, 2005)], and would greatly increase user productivity. Citation issues have previously been a hindrance to development of this type of server, but the unquestionable value of such a resource ensures a resolution. It was agreed that each tool contributor should be included as an author in the publication describing such a server.

Data formats

As mentioned previously, standardization of data formats is desirable for the epitope datasets as well as for the input and output of the diverse tools. We propose to develop a format suitable for the representation of B-cell epitope information which is flexible enough to be easily interpreted by computers and convertible into existing or user-specified formats. This would allow users to import B-cell epitope data into pre-existing programs for data analysis, as opposed to building applications *de novo*. For example, converting to a genome browser compatible format, such as General Feature Format (GFF) would allow for the visual comparison of multiple algorithms on a mature platform as well as the calculation of metrics commonly used in genome annotation such as enrichment values. This type of interoperability can easily be achieved with, for example, a well-developed XML schema.

Another key advantage of this format is that it is capable of capturing both sequence and structure information, including interacting residues. Several different structural definitions of what constitutes the antibody binding site in an antigen are used by different groups. Differing opinions exist regarding which residues in an antibody are relevant and what intermolecular distance is used to define an interaction. The solution for the time being is to use multiple definitions; however, these are issues that must be resolved in the future.

CONCLUSIONS

It is clear from the presentations and discussions at this workshop that the current state of B-cell epitope prediction is far from ideal and that the ultimate need of the user community is access to better prediction tools. However, in order to develop more accurate tools, it is imperative that the recommendations for assembling high-quality training and testing datasets be followed. Additionally, the periodic re-evaluation of each tool's performance on these datasets according to a common metric is vital. This approach will allow tool developers to draw upon the most effective aspects of each algorithm in order to evolve a superior one. The implementation of the recommendations developed at this workshop will greatly enhance the development of epitope prediction tools. We would hope that the grounded approach outlined here would garner the confidence of the immunology community and would provoke experimental researchers to employ these tools in their own work. Ultimately, this report should be considered a call for action and feedback from the greater community on the proposals drafted here. As mentioned previously, we will provide a web-based forum for this purpose (http://www.immuneepitope.org/jive/) and we explicitly welcome any contributions or collaborations as a result.

Acknowledgments

The authors would like to acknowledge the NIAID, in particular Dr Alison Deckhut Augustine for critical review of the manuscript and the Division of Allergy, Immunology and Transplantation for providing organizational and funding support for the workshop participants. Molecular graphics images were produced using the UCSF Chimera package (Pettersen *et al*., 2004) from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081). Kirin publication number 818.

REFERENCES

Batori V, Friis EP, Nielsen H, Roggen EL. 2006. An in silico method using an epitope motif database for predicting the location of antigenic determinants on proteins in a structural context. *J. Mol. Recognit.* **19**: 21–29.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28**: 235–242.

Blythe MJ, Flower DR. 2005. Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein Sci.* **14**: 246–248.

Bradford JR, Westhead DR. 2005. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics* **21**: 1487–1494.

Brusic V, Bajic VB, Petrovsky N. 2004. Computational methods for prediction of T-cell epitopes—a framework for modelling, testing, and applications. *Methods* **34**: 436–443.

Bujnicki JM, Elofsson A, Fischer D, Rychlewski L. 2001. Structure prediction meta server. *Bioinformatics* **17**: 750–751.

Cachau RE, Gonzalez-Sapienza G, Burt SK, Ventura ON. 2003. A new addition to the structural bioinformatics toolbox: 3D models of short bioactive peptides from multiple sequences using feedback restrained molecular dynamics (FRMD). *Cell Mol. Biol. (Noisy-le-grand)* **49**: 973–983.

Churchill ME, Stura EA, Pinilla C, Appel JR, Houghten RA, Kono DH, Balderas RS, Fieser GG, Schulze-Gahmen U, Wilson IA. 1994. Crystal structure of a peptide complex of anti-influenza

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/jmr

peptide antibody Fab 26/9. Comparison of two different antibodies bound to the same peptide antigen. *J. Mol. Biol.* **241**: 534–556.

Comeau SR, Gatchell DW, Vajda S, Camacho CJ. 2004. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **20**: 45–50.

Crowe JE Jr, Suara RO, Brock S, Kallewaard N, House F, Weitkamp JH. 2001. Genetic and structural determinants of virus neutralizing antibodies. *Immunol. Res.* **23**: 135–145.

De Groot AS. 2006. Immunomics: discovering new targets for vaccines and therapeutics. *Drug Discov. Today* **11**: 203–209.

Duhovny D, Nussinov R, Wolfson HJ. 2002. *Proceedings of the 2nd Workshop on Algorithms in Bioinformatics* (*WABI*), Guigo R, Gusfield D (eds). Springer Verlag: Rome; 185–200.

Haste Andersen P, Nielsen M, Lund O. 2006. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci.* **15**: 2558–2567.

Kawashima S, Ogata H, Kanehisa M. 1999. AAindex: Amino Acid Index Database. *Nucleic Acids Res.* **27**: 368–369.

Kernytsky A, Rost B. 2003. Static benchmarking of membrane helix predictions. *Nucleic Acids Res.* **31**: 3642–3644.

Koh IY, Eyrich VA, Marti-Renom MA, Przybylski D, Madhusudhan MS, Eswar N, Grana O, Pazos F, Valencia A, Sali A, Rost B. 2003. EVA: evaluation of protein structure prediction servers. *Nucleic Acids Res.* **31**: 3311–3315.

Korber BTM, Brander C, Haynes BF, Koup R, Moore JP, Walker BD, Watkins DI. 2005. HIV Molecular Immunology 2005. Los Alamos National Laboratory: Los Alamos.

Kulkarni-Kale U, Bhosle S, Kolaskar AS. 2005. CEP: a conformational epitope prediction server. *Nucleic Acids Res.* **33**: W168–W171.

Larsen JE, Lund O, Nielsen M. 2006. Improved method for predicting linear B-cell epitopes. *Immunome Res.* **2**: 2.

Lund O, Nielsen M, Lundegaard C, Kesmir C, Brunak S. 2005. Immunological Bioinformatics. The MIT Press: Cambridge.

Meloen RH, Puijk WC, Slootstra JW. 2000. Mimotopes: realization of an unlikely concept. *J. Mol. Recognit.* **13**: 352–359.

Mumey BM, Bailey BW, Kirkpatrick B, Jesaitis AJ, Angel T, Dratz EA. 2003. A new method for mapping discontinuous antibody epitopes to reveal structural features of proteins. *J. Comput. Biol.* **10**: 555–567.

Mumey B, Ohler N, Angel T, Jesaitis A, Dratz E. 2006. *Proceedings of High Performance Computing in Genomics, Proteomics, and Transcriptomics*. Springer-Verlag: Sorrento.

Neuvirth H, Raz R, Schreiber G. 2004. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J. Mol. Biol.* **338**: 181–199.

Odorico M, Pellequer JL. 2003. BEPITOPE: predicting the location of continuous epitopes and patterns in proteins. *J. Mol. Recognit.* **16**: 20–22.

Parker JM, Guo D, Hodges RS. 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* **25**: 5425–5432.

Pellequer JL, Westhof E, van Regenmortel MH. 1993. Correlation between the location of antigenic sites and the prediction of turns in proteins. *Immunol. Lett.* **36**: 83–99.

Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, Fleri W, Kronenberg M, Kubo R, Lund O, Nemazee D, Ponomarenko JV, Sathiamurthy M, Schoenberger S, Stewart S, Surko P, Way S, Wilson S, Sette A. 2005. The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol.* **3**: e91.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**: 1605–1612.

Saha S, Bhasin M, Raghava GP. 2005. Bcipep: a database of B-cell epitopes. *BMC Genomics* **6**: 79.

Saha S, Raghava GP. 2006. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* **65**: 40–48.

Saini HK, Fischer D. 2005. Meta-DP: domain prediction meta-server. *Bioinformatics* **21**: 2917–2920.

Schlessinger A, Ofran Y, Yachdav G, Rost B. 2006. Epitome: database of structure-inferred antigenic epitopes. *Nucleic Acids Res.* **34**: D777–D780.

Smith GP, Petrenko VA. 1997. Phage display. *Chem. Rev.* **97**: 391–410.

Sollner J. 2006. Selection and combination of machine learning classifiers for prediction of linear B-cell epitopes on proteins. *J. Mol. Recognit.* **19**: 209–214.

Sollner J, Mayer B. 2006. Machine learning approaches for prediction of linear B-cell epitopes on proteins. *J. Mol. Recognit.* **19**: 200–208.

Swets JA. 1988. Measuring the accuracy of diagnostic systems. *Science* **240**: 1285–1293.

Toseland CP, Clayton DJ, McSparron H, Hemsley SL, Blythe MJ, Paine K, Doytchinova IA, Guan P, Hattotuwagama CK, Flower DR. 2005. AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res.* **1**: 4.

Underwood PA. 1984. An antigenic map of the haemagglutinin of the influenza Hong Kong subtype (H3N2), constructed using mouse monoclonal antibodies. *Mol. Immunol.* **21**: 663–671.

van Regenmortel MH. 2001. Antigenicity and immunogenicity of synthetic peptides. *Biologicals* **29**: 209–213.

van Regenmortel MH. 2006. Immunoinformatics may lead to a reappraisal of the nature of B cell epitopes and of the feasibility of synthetic peptide vaccines. *J. Mol. Recognit.* **19**: 183–187.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/jmr