Immunology
The Journal of cells, molecules, systems and technologies

Immunoinformatics: an integrated scenario

Namrata Tomar and Rajat K. De
Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India

Summary

Genome sequencing of humans and other organisms has led to the accumulation of huge amounts of data, which include immunologically relevant data. A large volume of clinical data has been deposited in several immunological databases and as a result immunoinformatics has emerged as an important field which acts as an intersection between experimental immunology and computational approaches. It not only helps in dealing with the huge amount of data but also plays a role in defining new hypotheses related to immune responses. This article reviews classical immunology, different databases and prediction tools. It also describes applications of immunoinformatics in designing *in silico* vaccination and immune system modelling. All these efforts save time and reduce cost.

Keywords: allergy; B cells; *in silico* models; major histocompatibility complex/human leucocyte antigen; T cells

Introduction

The term 'immunity' was developed to describe individuals who had recovered from certain infectious diseases and were protected from the same diseases when they were re-encountered. An immune system and associated biological processes exist within these individuals, which are responsible for developing 'immunity'. The role of an immune system is to protect against diseases by identifying and killing pathogens. An immune system includes innate and adaptive components. According to the traditional dogma of immunology, vertebrates have both innate and adaptive immune systems whereas invertebrates possess only an innate immune system.[1] The innate immune system acts more rapidly, and is older and more evolutionarily conserved than the adaptive immune system. It provides the backbone on which the adaptive immune system was able to evolve. The innate immune system is less specific and works as a first line of defence.[2] It comprises four types of defensive barriers, namely, anatomic (e.g. skin and mucous membranes), physiological (e.g. temperature, low pH), phagocytic (e.g. blood monocytes, neutrophils, tissue macrophages) and inflammatory (e.g. serum proteins). An adaptive immune response occurs against a pathogen within 5 or 6 days after the initial exposure to the pathogen.[2] It has evolved in vertebrates as a defence system. Functionally, it accounts for two inter-related activities: recognition and response. It can discriminate between the body's own cells and proteins from foreign molecules, and can recognize chemical differences between two pathogens. It can also recognize altered self cells, such as virus-infected self cells, and distinguish between healthy and cancerous cells. However, it may not always recognize cancer cells as foreign or abnormal cells. As soon as the adaptive immune system recognizes a pathogen, an effector response is elicited to kill or neutralize it. The response is unique to defend against a particular type of pathogen. Later exposure to the same pathogen induces a heightened and more specific response because the adaptive immune system retains memory.

The adaptive immune system has two parts: the cellular immune response of T cells and the humoral response of B cells.[2,3] An antigen has a specific small part, known as the epitope, which is recognized by the corresponding receptor present on B or T cells. B-cell epitopes can be linear and discontinuous amino acids. T-cell epitopes are short linear peptides. Most of the T cells can be in either of the two subsets, distinguished by the presence of one or other of two glycoproteins on their surface, designated as CD8 or CD4. CD4 T cells function as T helper (Th) cells that recognize peptides displayed by major histocompatibility complex (MHC) class II molecules. On the other hand, CD8 T cells function as cytotoxic T (Tc) cells, which recognize peptides displayed by MHC class I molecules. A brief description of various components of the human immune system is provided as supplementary material. The idea that the immune response exists in an organism is quite old. The earliest literary reference to immunology goes back to 430 BC by Thucydides.[2] In 1798, Edward Jenner found some milkmaids who were immune to smallpox because they had earlier contracted cowpox (a mild disease). The next major advancement in immunology came with the induction of immunity to cholera by Louis Pasteur. After applying weakened pathogen to animals, he administered (in 1885) a dose of vaccine to a boy bitten by a rabid dog and the boy survived. However, Pasteur could not explain its mechanism. In 1890, experiments of Emil Von Behring and Shibasabura Kitasato led to the understanding of the mechanism of immunity. Their experiments described how antibodies present in the serum provided protection against pathogens.

An immune system may be considered as a network of thousands of molecules, which leads to many intertwined responses. It is structurally and functionally diverse and this diversity varies both between individuals and temporally within individuals. Huge amounts of data related to immune systems are being generated. Immunologists have been using high throughput experimental techniques for a long time, which have generated a vast amount of functional, clinical and epidemiological data. The development of new computational approaches to store and analyse these data are needed. Recently, immunology-focused resources and software are appearing, which help in understanding the properties of the whole immune system.[4] This has given rise to a new field, called immunoinformatics. Immunogenomics, immunoproteomics, epitope prediction and *in silico* vaccination are different areas of computational immunological research. Recently, Systems Biology approaches have been applied to investigate the properties of the dynamic behaviour of an immune system network.

Immunoinformatics includes the study and design of algorithms for mapping potential B- and T-cell epitopes, which lessens the time and cost required for laboratory analysis of pathogen gene products. Using this information, an immunologist can explore the potential binding sites, which, in turn, leads to the development of new vaccines. This methodology is termed 'reverse vaccinology' and it analyses the pathogen genome to identify potential antigenic proteins.[5] This is advantageous because conventional methods need to cultivate pathogen and then extract its antigenic proteins. Although pathogens grow fast, extraction of their proteins and then testing of those proteins on a large scale is expensive and

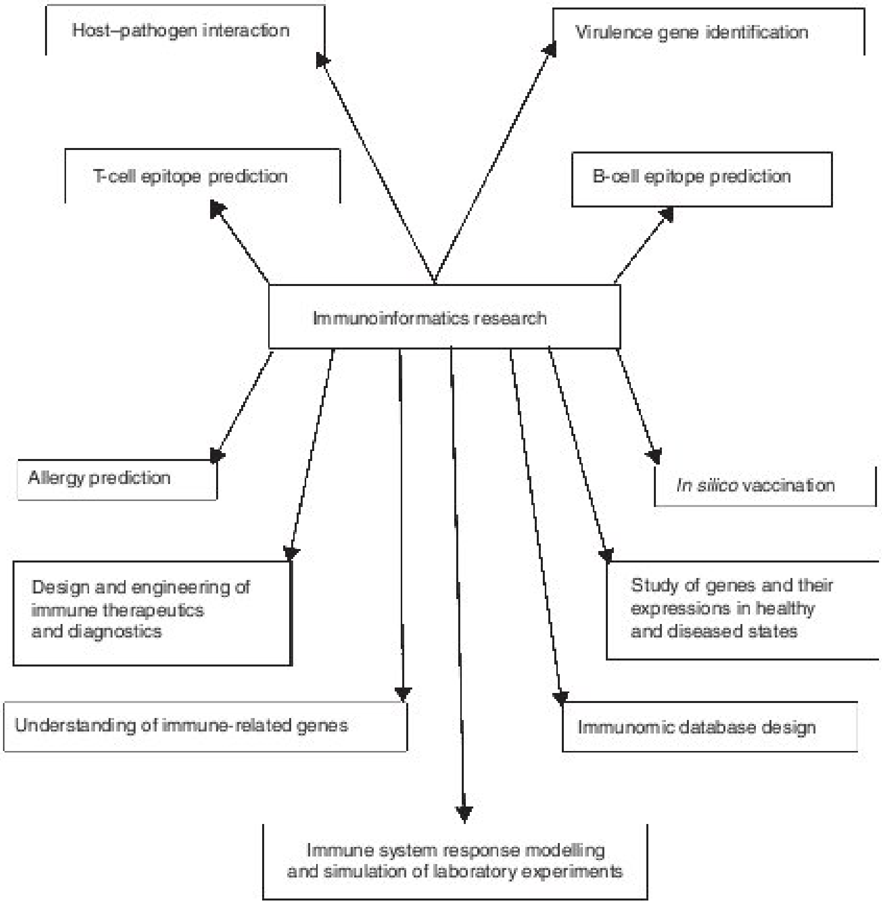
Host–pathogen interaction
Virulence gene identification
T-cell epitope prediction
B-cell epitope prediction
Immunoinformatics research
Allergy prediction
In silico vaccination
Design and engineering of immune therapeutics and diagnostics
Study of genes and their expressions in healthy and diseased states
Understanding of immune-related genes
Immunomic database design
Immune system response modelling and simulation of laboratory experiments

Figure 1. Immunoinformatics: research areas.

time consuming. Immunoinformatics is capable of identifying virulence genes and surface-associated proteins.

Figure 1 shows the different research areas of immunoinformatics. All of these areas are described in separate sections of this article. We describe various available information regarding classical immunology, different immunomic databases, and B-cell and T-cell epitope prediction tools and softwares. Several methods are now available that enable one to map epitopes and design therapeutic vaccines more quickly. Some of them are described in this article, which concludes with some applications of immunoinformatics.

Immunomics

The term 'immunome' corresponds to all the genes and proteins taking part in immune responses. It excludes genes and proteins that are expressed in cell types other than in immune cells.[6] According to Sette *et al.*[7] all immune reactions that are the result of interactions between the host and antigenic peptides are referred to as 'immunome reactions', and their study is entitled 'immunomics'. Like genomics and proteomics, immunomics is a new discipline that uses high throughput techniques to understand the immune system mechanism.[8,9]

Various datatypes and databases

In this section, we focus on various immune-system-related datatypes and databases. A brief description of these databases is provided. The section starts with some experimental techniques and results.

Experimental data

There has been an explosion in available experimental data in immunology as the result of the advent of high throughput molecular biology techniques. These techniques help in finding the structure and function of immune genes and their products.[10]

There are many immunological techniques that are used to understand the underlying mechanism of an immune system and its responses to various infections, diseases and drug administration, namely, affinity chromatography,[11] flow cytometry,[12] radioimmunoassay,[13] enzyme-linked immunosorbent assay,[13,14] competitive inhibition assay[15] and Coombs test.[16] Here, we present some experimental findings, which help to identify B-cell and T-cell epitopes and to study immune responses.

Experimental techniques for exploring immune system components The ability to identify epitopes in the immune response has important implications in the diagnosis of diseases. For this reason, epitopes for B and T cells need to be identified and mapped. In this context, Wanga *et al.*[17] mapped the B-cell epitopes present on non-structural protein 1 (NS1), i.e. NS1-18 and NS1-19 in Japanese encephalitis virus. For epitope mapping, a series of 51 partially overlapping fragments covering the entire NS1 protein were expressed with a glutathione S-transferase tag and then screened with a monoclonal antibody. They found that the motif of (146) EHARW (150) was the minimal unit of the linear epitope recognized by that monoclonal antibody.

Purification techniques like affinity chromatography are used to purify MHC–peptide from membrane MHC molecules, which can be analysed by capillary high-pressure liquid chromatography electrospray ionization-tandem mass spectrometry.[18] They can be further used to find new tumour-associated antigens. These are proteins that are not unique to cancer cells but are expressed in tumour cells. One approach to find tumour-associated antigens is based on transfection of the expression library made from complementary DNA into cells expressing the desired MHC haplotypes.[19] The clones are selected on the basis of their ability to provoke an immune response in T cells of the individuals with the same MHC type. MHC–peptide complexes are required for tumour therapeutics.

Dengue, a human viral disease transmitted by arthropod vectors, has an annual mortality rate of 25 000.[20] Dengue fever and dengue haemorrhagic fever are caused by the four dengue viruses, DEN-1, -2, -3 and -4, which are closely related antigenically. Random Peptide Libraries of peptides displayed on the phage help in selecting sequences that mimic epitopes from microorganisms. Amin *et al.*[21] used Random Peptide Libraries and identified two peptides, NS3 and NS4B. These two non-structural proteins resemble the antigenic structure of B-cell epitopes of dengue virus obtained from a phage-peptide library using human polyclonal antisera from patients who had recovered from dengue virus infection. These two peptides could be used for the development of a diagnostic kit and a potential vaccine.

Immunomic microarray technology and analysis

Using DNA microarray technology, one can measure the RNA expression of thousands of genes simultaneously in a single assay. The principle of all kinds of microarray technologies is binding and measurement of target biological specimens to complementary probes. Similar technology is used in functional immunomics and is referred to as 'immunomic microarray'. It includes dissociable antibody microarray,[22] serum microarray[23] and serological analysis of a complementary DNA expression library (SEREX).[24]

An antibody microarray consists of antibody probes and antigen targets, so that it can be used to measure concentration of antigen for a specific antibody probe, but peptide microarray has an opposite approach. It uses antigen peptides as fixed probes and serum antibodies as targets. The recent technology is peptide–MHC microarray or the artificial antigen-presenting chip. In this technique, recombinant peptide–MHC complexes and co-stimulatory molecules are immobilized on a surface, and a population of T cells is incubated with the microarray. The T-cell spots act as artificial antigen-presenting cells[25] containing defined MHC-restricted peptides. The advantage of using peptide–MHC is that it can map the MHC-restricted T-cell epitope.

The proteins responsible for the normal functioning of the cellular machinery may have sequence similarity with various pathogenic microbes. They can induce autoimmunity and thereby are less useful for vaccine development. Microarray technology helps in selecting these proteins from genomic sequences.[26] It is being applied in autoimmune disease diagnosis and treatment,[27] allergy prediction,[28] T-cell and B-cell epitope mapping[29] and vaccination.[30] The immunomic and genomic microarray data differ in several ways, e.g. both of them have different designs. One can measure two or more signals simultaneously determined by a single feature, i.e. epitope in immunomic microarray.[31,32] DNA microarrays measure one response value for each gene per sample, i.e. messenger RNA concentration produced by the gene, but a single epitope can generate different response values corresponding to different epitopes in peptide–MHC chips. In the case of the B-cell epitope, it can be recognized by different isotypes of immunoglobulins, so here, one can measure both intensity and quality of the antibody response.

Immunomic databases

Knowledge of B-cell and T-cell epitope-mediated responses has been increased dramatically. Epitope information-related databases, bioinformatics tools and prediction algorithms help in understanding the structure and sequences of amino acids of epitopes. This knowledge is crucial for basic immunological studies, diagnosis and treatment of various diseases, and in vaccine research.[33] InnateDB[34] (http://www.innatedb.ca) has been created to understand the complete network of pathways and interactions of innate immune system responses. It is an integrated biological database of the human and mouse molecules with 100 000 experimentally verified interactions and 2500 pathways involved in innate immunity. It has a newer version, called Cerebral,[35] which is a Java plugin for the Cytoscape biomolecular interaction viewer[36] for automatically generating layouts of biological pathways. Table 1 lists some of the databases that deal with information related to B-cell epitopes, T-cell epitopes, allergy prediction and evolution of immune system genes and proteins.

B-cell epitope databases

Conformational epitopes have implicit structural information related to antigen and binding mode. It has been found that 90% of B-cell epitopes are conformational or discontinuous. Bcipep[37] (http://www.imtech.res.in/raghava/bcipep) provides comprehensive information about experimentally verified B-cell epitopes and tools for mapping these epitopes on an antigen sequence. Immunogenicity of a peptide in Bcipep is divided into three dimensions: immunodominant, immunogenic and null immunogenic. Searches can be restricted to the basis of immunogenicity. Bcipep has some limitations such as, (i) it contains no discontinuous epitopes, (ii) it includes a limited number of unique peptides, and (iii) it provides information on

Table 1. Databases on B-cell epitopes, T-cell epitopes, allergen and molecular evolution of immune system components

Databases	Names	URLs	References
B-cell epitopes	CED	http://immunet.cn/ced/	[38]
	Bcipep	http://www.imtech.res.in/raghava/bcipep	[37]
	Epitome	http://cubic.bioc.columbia.edu/services/epitome/	[39]
	IEDB	http://www.immuneepitope.org/	[33]
	IMGT®	http://imgt.cines.fr	[43]
T-cell epitopes	JenPep	http://www.darrenflower.info/jenpep/	[40]
	Syfpeithi	http://www.syfpeithi.de	[41]
	IEDB	http://www.immuneepitope.org/	[33]
	FRED	http://www-bs.informatik.uni-tuebingen.de/Software/FRED	[42]
	IMGT®	http://imgt.cines.fr	[43]
Allergen	Database of IUIS	http://www.allergen.org	[47]
	AllergenPro	http://www.niab.go.kr/nabic/	[48]
	SDAP	http://fermi.utmb.edu/SDAP/	[49]
Information related to molecular evolution of immune system components	ImmTree	http://bioinf.uta.fi/ImmTree	[50]
	Immunome database	http://bioinf.uta.fi/Immunome/	[51]
	ImmunomeBase	http://bioinf.uta.fi/ImmunomeBase	[52]
	Immunome Knowledge Base	http://bioinf.uta.fi/IKB/	[6]

peptides containing only natural amino acids. CED[38] (Conformational Epitope Database) can be used for the evaluation and improvement of existing epitope prediction methods. CED 0.03 release (http://immunet.cn/ced/) has 293 entries. It has a collection of B-cell epitopes from the literature, conformational epitopes defined by methods like X-ray diffraction, nuclear magnetic resonance, scanning mutagenesis, overlapping peptides and phage display. CED maintains well-defined conformational epitope information. It rejects conformational epitopes that are not defined clearly so the database is small. Epitome[39] (http://cubic.bioc.columbia.edu/services/epitome/) contains all known antigen–antibody complex structures. A semi-automated tool has also been developed that identifies the antigenic interactions within the known antigen–antibody complex structures and compiled these interactions into Epitome. None of the other databases can locate the complementary determining regions or identify the antigenic residues semi-automatically. Epitome updating follows the updating of SCOP, i.e. Epitome is updated twice a year, as soon as SCOP is updated.

If we compare Epitome and CED, we find that they are similar in size, the difference lies in the source of collection of B-cell epitopes. Epitome collects B-cell epitopes only from Protein Data Bank (PDB) structures and includes information on complementary determining regions. In contrast, CED takes data from the literature and from the above-mentioned methods. As their sources are different, one can use the complementary information.

T-cell epitope databases

T-cell epitopes do not always have high affinity for MHC binders. A functional T-cell response requires MHC–peptide binding and a proper interaction of the MHC–peptide ligand with a specific T-cell receptor (TR). We need well-characterized data to model the process of binding of peptides to transfer associated protein (TAP) and MHCs, which function as T-cell epitopes. Some recent investigations include finding and mapping of potential epitopes. Epitope mapping leads to the design of effective vaccines. JenPep[40] (latest updated version 2.0) (http://www.darrenflower.info/jenpep/) is a relational database with five types of data: a compilation of quantitative measures of binding for 12 336 entries of peptides to MHC I and II, an annotated list of 3218 entries of dominant and subdominant T-cell epitopes, and a set of over 441 records of quantitative data for peptide binding to TAP peptide transporter. In the latest update (i.e. in version 2.0), two new categories have been introduced: B-cell epitopes (816 entries) and peptide–MHC–TR complex formation (49 entries).

The Syfpeithi database[41] (http://www.syfpeithi.de) has information on MHC class I and II anchor motifs, and binding specificity. It calculates a score based on the following rules: calculated score values differentiate among anchor, auxiliary anchor or preferred residues.

FRED[42] (http://www-bs.informatik.uni-tuebingen.de/Software/FRED) deals with methods of data processing. It also compares the performance of prediction methods by considering experimental values. It can handle polymorphic sequences. IMGT®[43] (the international ImMunoGeneTics information system®; http://imgt.cines.fr) has a good collection of immunoglobin, T-cell receptor, MHC, and related proteins of the immune system of humans and other vertebrates. It has five databases and 15 interactive online tools for sequence, genome and three-dimensional structural analysis.

IEDB 2.0,[33] (Immune Epitope Database and Analysis Resource Database) (http://www.immuneepitope.org/), sponsored by the National Institute for Allergy and Infectious Diseases (http://www.niaid.nih.gov), has different tools to find B-cell and T-cell epitopes. It contained details of 75 056 peptide epitopes till July 2010.

It also facilitates the conversion of experimental data from text and figures in a journal publication into a computer-friendly format in the form of ONTIEs (Ontology of Immune Epitopes) (http://ontology.iedb.org). This module has been imported by the OBI (Ontology for Biomedical Investigations) Consortium (http://purl.obolibrary.org/obo/obi).[44]

Allergy prediction databases

Allergens are proteins or glycoproteins recognized by immunoglobulin E (IgE), which is produced by the immune system in allergic individuals. So far, 1500 allergenic structures have been identified.[45] Online allergen databases and allergy prediction tools are being used to find cross-reactivity between known allergens. Localization of B and T cells in the allergen may not coincide.[46] The differences between both kinds of epitopes present in an antigen are: T-cell epitopes are only linear (as mentioned earlier) and are distributed throughout the primary structure of the allergen, whereas B-cell epitopes can be either linear or conformational, recognized by IgE antibodies, and are located on the surface of the molecule accessible to antibodies. Moreover, in the case of B-cell epitopes, predicting allergenicity in a molecule based on known conformational epitopes is a difficult task.

The Allergen Nomenclature database of the International Union of Immunological Societies (IUIS) has an allergen database[47] (http://www/allergen.org). The Allergen Pro database[48] (http://www.niab.go.kr/nabic/) contains information related to 2434 allergens, e.g. allergens in rice microbes (712 records), animals (617 records) and plants (1105 records). The web server Allergome 4.0[45] (http://www.allergome.org) provides an exhaustive repository of IgE-binding compound data. It has a total 1736 allergen sources (updated in March 2010). The Real-Time Monitoring of IgE sensitization module (ReTiME), in Allergome 4.0, enables one to upload raw data from both *in vivo* and *in vitro* experiments. This is the first attempt where information technology has been applied to allergy data mining. SDAP[49] (Structural database of Allergenic Proteins) (http://fermi.utmb.edu/SDAP/) is a web server that provides cross-referenced access to the sequence and structure of the IgE epitope of allergenic proteins. Its algorithm is based on conserved properties of amino acid side chains. In its latest update, it has 887 allergenic proteins.

Databases related to molecular evolution of immune genes and proteins

To explore the molecular evolution of the human immune system, a reference set of genes and proteins must be defined. For this reason, Ortutay *et al.*[50] constructed a database ImmTree (http://bioinf.uta.fi/ImmTree) for the evolutionary trees of proteins of the human immune system. It contains information for orthologues of the human genes in 80 species. The Immunome database[51] (http://bioinf.uta.fi/Immunome/) is another database in which 847 genes and proteins are annotated and characterized according to their functions, protein domains and gene ontology terms from the human immunome.

A vast amount of molecular data for genes and proteins for the immune system has accumulated. The Immunome Knowledge Base (IKB)[6] is a single service access to many immune system databases and resources. It combines the other databases, namely Immunome[51] and ImmunomeBase,[52] and several additional data items in an integrated fashion. It has orthologue groups of 1811 metazoan immunity genes for studying the evolution of the immune system, and includes the evolutionary history of genes and proteins, orthologous genes, information on disease-causing mutations, alternatively spliced variants and copy number variations.

Various tools and algorithms

Here, we throw some light on available immunology-related tools and algorithms. Traditionally, determination of the binding affinity of MHC molecules and antigenic peptides is the main objective when predicting epitopes. The experimental techniques are found to be difficult and time consuming. As a result, several *in silico* methodologies are being developed and used to identify epitopes. These techniques include matrix-driven methods, finding structural binding motifs, a quantitative structure–activity relationship (QSAR) analysis, homology modelling, protein threading, docking techniques and design of several machine-learning algorithms and tools. In the past, computational techniques could only identify sequence characteristics but new improved algorithms and tools are being designed to increase the predictive performance. Table 2 lists some of the tools that deal with B-cell and T-cell epitope prediction, allergy prediction and *in silico* vaccination. Here, we describe different methodologies for epitope and allergy prediction, and the process of *in silico* vaccination briefly.

B-cell epitope prediction

B cells produce antibodies that are protein in nature. B-cell epitopes are antigenic determinants on the surface of pathogens that interact with B-cell receptors. The B-cell receptor binding site is hydrophobic with six hypervariable loops of variable length and amino acid composition. As described in ref.[53], B-cell epitopes are classified as continuous/linear and discontinuous/conformational. Most of the B-cell epitopes are discontinuous where distant residues are brought into spatial proximity by protein folding. Experiments are mostly based on linear epitopes. There are both sequence-based and structure-based prediction tools but prediction tools are limited for discontinuous B-cell epitopes.[37,54]

Prediction using amino acid propensity scale

Classically, amino acid propensity scales such as hydrophilicity and characteristic flexibility have been used to identify epitopes present in antigens. Pellequer *et al.*[55] compared several propensity scale methods using a dataset of 14 epitope annotated proteins and found that the scales of Parker *et al.*[56] Chou and Fasman,[57] Levitt,[58] and Emini *et al.*[59] provide better results than the other scales tested.[53] El-Manzalawy *et al.*[60] compared propensity-scale-based methods with a Naive Bayes classifier. They used three different representations of the classifier input: amino acid identities, position-specific scoring matrix profiles and dipeptide composition. They used two datasets, one is the propensity dataset and the other is from BciPep.[37] They considered 125 non-redundant antigens at 30% sequence similarity cut off from BciPep. The Bepitope tool[61] predicts continuous epitopes based on the prediction of protein turns. It is a newer version of PREDITOP[62] and uses more than 30 propensity scale values. The Bcepred server[63] (http://www.imtech.res.in/raghava/bcepred/) predicts linear B-cell epitopes with 58·7% accuracy based on combined amino acid properties like accessibility, hydrophilicity, flexibility, polarity, exposed surface and turns.

Analyses of antigen–antibody interactions are performed on antibody-binding sites on proteins, which help in predicting the linear and conformational B-cell epitopes. Taking this into consideration, a database,

Conclusions

There have been several computational methods developed to predict miRNAs. Five different methods based on two different categories of miRNA gene identification tools have been compared to understand their relative performance. Among all the tools, MiPred shows the best performance. One class approach can be a good alternative but as far as overall accuracy is concerned, certain improvements need to be incorporated for a better performance. Moreover, BayesSVMmiRNAfind using SVM and naïve Bayes classifier show lowest specificity although the sensitivity is quite high in both the cases.

References

1. Adai A, Johnson C, Mlotshwa S, Evans SA, Manocha V, et al. (2005) Computational prediction of miRNAs in Arabidopsis thaliana. Genome Research, 15: 78-91.

2. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281-297.

3. Bentwich I (2005) Prediction and validation of microRNAs and their targets. FEBS Lett 579: 5904 - 5910.

4. Berezikov E, Guryev V, Belt JV, Weinholds E, Plasterk RHA, et al. (2005) Phylogenetic Shadowing and Computational Identification Of Human microRNA genes. Cell 120: 21-24.

5. Bonnet E, Wuyts J, Rouze P, Peer YV (2004) Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identify important target genes. PNAS 101: 11511-11516.

6. Brameier M, Wiuf C (2007) *Ab initio* identification of human microRNAs based on structure motifs. BMC Bioinformatics 8: 478.

7. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nature Genetics 38: 813-818.

8. Dezulian T, Remmert M, Palatnik JF, Weigel D, Huson DH (2006) Identification of plant microRNA homologs. Bioinformatics 22: 359-360.

9. Dror G, Sorek R, Shamir R (2005) Accurate identification of alternatively spliced exons using support vector machine. Bioinformatics 21: 897-901.

10. Feng J, Sun G, Yan J, Noltner K, Li W, et al. (2009) Evidence for X-chromosomal schizophrenia assoaciated with miRNA alterations. PLoS ONE 4: e6121.

11. Frank E, Hall M, Trigg L, Holmes G, Witten I (2004) Data mining in bioinformatics using Weka. Bioinformatics 2479-2481

12. Gard OS, Thomassen, Rosok O, Rognes T (2006) Computational Prediction of MicroRNAs Encoded in Viral and Other Genomes. Journal of Biomedicine and Biotechnology Article ID 95270: 1-10.

13. Griffiths-Jones S, Saini HK, Van DS, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acid Research 36: D154-D158.

14. Griffiths-Jones S, Grocock RJ, Van DS, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acid Research 34: D140-D144.

15. Griffiths-Jones S (2004) The microRNA Registry. Nucleic Acid Research 32: D109-D111.

16. Gruber AR, Neubock R, Hofacker IL, Washietl S (2007) The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures. Nucleic Acids Research 35: W335-W338.

17. Helvik SA, Snove O, Saetrom P (2007) Reliable prediction of Drosha processing sites improves microRNA gene prediction. Bioinformatics Genome Analysis 23: 142-149.

18. Hertel JA, Stadler PF (2006) Hairpins in a Haystack: recognizing microRNA precursors in comparative genomics data. Bioinformatics 22: e197-e202

19. Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Research 31: 3429-3434

20. Jiang P, Wu H, Wang W, Wa M, Sun X, et al. (2007) MiPred: classification of real and pseudo microRNA precursor using random forest prediction model with combined features. Nucleic acid research, 35: W339-W344.

21. Jones SG (2004) The microRNA Registry. Nucleic Acids Research 32: D109-D111.

22. Lai EC, Tomancak P, Williams RW, Rubin GM (2003) Computational identification of *Drosophila* microRNA genes. Genome Biology 4: R42.

23. Lee RC, Feinbaum RL, Ambros V (1993) The C.

and Los Alamos human immunodeficiency virus (HIV) database (http://www.hiv.lanl.gov). They tested a number of propensity scale methods on the Pellequer *et al.* dataset,[55] and found the best scale to be by Levitt.[58] Then, they used a Hidden Markov model (HMM) to predict the location of linear B-cell epitopes and tested HMMs on the Pellequer *et al.* dataset to find optimal parameters. HMM was combined with one set of the two best propensity scale methods, i.e. Parker *et al.*[56] and Levitt[58] to get more accurate predictions.

Prediction methodology for discontinuous B-cell epitopes

As mentioned earlier, more than 90% of B-cell epitopes are discontinuous but they may comprise a linear amino acid chain of peptides, which is brought closure in three-dimensional space.[69] There is a specialized form of protein–protein interaction in these epitopes. Changes in protein folding may lead to changes in the number of epitopes.[46] The characterization and prediction of B-cell epitopes are mainly conformation dependent so the task of prediction is more difficult compared with that of T-cell epitopes. The most accurate way to identify the B-cell epitope is through X-ray crystallography. Anderson *et al.*[70] presented a method called Discotope, (http://www.cbs.dtu.dk/services/DiscoTope/), which is a combination of amino acid statistics, spatial information and surface exposure. It was trained on a dataset of discontinuous epitopes of 76 X-ray structures of antibody–antigen complexes. It detects 15·5% of residues located in discontinuous epitopes with a specificity of 95%. The conventional Parker hydrophilicity scale (for predicting linear B-cell epitopes) identifies only 11·0% of residues with 95% specificity. It is said to be the first method developed for prediction of discontinuous B-cell epitopes with better performance than methods based only on sequence data.

Bublil *et al.*[71] developed Mapitope for conformational B-cell epitope mapping. The hypothesis behind Mapitope is that the simplest meaningful fragment of an epitope is an amino acid pair of residues that lie within the epitope, which are the result of folding. A set of affinity isolated peptides was obtained by screening the phage display peptide libraries with the antibody of interest. This set was given as algorithm input, and one to three epitope candidates on the surface of the atomic structure of the antigens were obtained as output.

A computational method has been presented by Sollner *et al.*[72] to automatically select and rank peptides for the stimulation of otherwise functionally altered antibodies. They investigated the integration of B-cell epitope prediction with the variability of antigen, and the conservation of patterns for posttranslational modification prediction. By their observation, they found high antigenicity, low variability and low likelihood of posttranslational modification for the identification of biorelevant sites. Greenbaum *et al.*[53] assembled non-redundant datasets of repetitive three-dimensional structure of antigen and antigen–antibody complexes from the PDB. The CEP web interface[73] (http://bioinfo.ernet.in/cep.htm) predicts conformational and sequential epitopes, and also antigenic determinants. It uses structure-based approaches, solvent accessibility of amino acids and spatial distance cut-off to predict antigenic determinants. Less availability of the three-dimensional structure data of protein antigens limits the utility of this server.

Mimotope-based epitope prediction methodology

Phage display library has a large number (more than 109) of random peptides.[74] It is widely used for finding protein–protein interactions (especially in antibody–antigen interactions), protein function identification and in development of new drugs and vaccines. These libraries are screened to find the pool of peptides that can bind to desired antibody. These pools of peptides are called mimotopes.[69,74,75] Mimotopes and antigens are both recognized by the same antibody paratope. Mimotopes are said to be the imitated part of the epitope. So, it is possible that a mimotope may have some valuable information about the epitope. However, homology may not exist between the mimotope and the epitope of the native antigen. This mimicry exists because of similarities in physiochemical properties and spatial organization.[75] Considering these properties, mimotope pools are used to mine information to predict an epitope. Using this concept, the MIMOP tool[75] has been developed. MIMOP predicts linear and conformational epitopes based on two algorithms: MimAlign uses degenerated alignment analyses, and MimCons is based on consensus identification. MIMOX[76] (http://web.kuicr.kyoto-u.ac.jp/~hjian/mimox) comes in the same category, which maps a single mimotope or a consensus sequence of a set of mimotopes onto the corresponding antigen structure. Then, it searches for all of the clusters of residues that could be the native epitope. Pepitope[74] (http://pepitope.tau.ac.il/) (an advanced server for mimotope-based epitope prediction approaches) uses two algorithms: Pepsurf[77] and Mapitope.[71] It maps each mimotope so as to map them onto the solved structure of the antigen surface. Alignment of the mimotope is done first in MIMOX; this step is different in Pepitope. If we compare it with MIMOP, MIMOP aligns the peptides to the antigen at the sequence level rather than directly to the three-dimensional structure. The three-dimensional structure is considered only after the alignment stage.

Sometimes linear peptides mimic conformational epitopes. The same phage display peptide libraries for screening with the respective antibodies are used to select these mimotopes. Schreiber *et al.*[78] presented a software, 3DEX (3D-Epitope-Explorer) (http://www.schreiber-abc.

com/3dex/) that allows localizing of linear peptide sequences within three-dimensional structures of proteins. Its algorithm takes into account the physiochemical neighbourhood of C-α or C-β atoms of individual amino acids and surface exposure of the amino acids. Authors were able to localize mimotopes from the plasma of patients who were HIV-positive within the three-dimensional structure of gp120. The epitopes defined by 3DEX are not proven by mathematical calculations and energy minimizations.

T-cell epitope prediction

It is necessary to bind antigenic peptides with MHC so that cytotoxic T cells can recognize them. Hence, identification of MHC binding peptides is a central part of any algorithm that predicts T-cell epitopes. There exist several methodologies for the prediction of MHC binding peptides, which are based on the idea of quantitative matrices, HMM, ANN, SVM and structure of the peptides.

Prediction through matrix-driven methods

Huang and Dai[79] first investigated a new encoding scheme of peptides. This scheme used the BLOSUM matrix with the amino acid indicator vectors for direct prediction of T-cell epitopes. It replaced each non-zero entry in the amino acid indicator vector by the corresponding value appearing in the diagonal entries in the BLOSUM matrix. The MMBPred[80] (http://www.imtech.res.in/raghava/mmbpred/) server predicts the mutated promiscuous and high-affinity MHC binding peptide. It uses the matrix data in a linear prediction model and ignores peptide conformation. The prediction is based on the quantitative matrices of 47 MHC alleles.

Prediction through HMM

Transfer Associated Protein is an important component of the MHC I antigen-processing and presentation pathway. A TAP transporter can translocate peptides of 8–40 amino acids into endoplasmic reticulum. Zhang *et al.*[81] developed PRED$^{\mathrm{TAP}}$ (http://antigen.i2r.a-star.edu.sg/predTAP) for the prediction of peptide binding to hTAP. They used a three-layer back propagation network with the sigmoid activation function. The inputs were the binary strings, representing nonamer peptide. Second, they used second-order HMM. The results are both sensitive and specific.

Prediction through ANN

Neilsen *et al.*[82] described an improved neural network model to predict T-cell class I epitopes. They have a combination of sparse encoding, BLOSUM encoding and input derived from HMM. The dataset consists of 528 nonamer amino acid peptides for which the binding affinity to the HLA I molecule A*0204 has been measured in a method described by Buus *et al.*[83] NetCTL server[84] (http://www.cbs.dtu.dk/services/NetCTL/) uses a method to integrate the prediction of peptide MHC class I binding, proteasomal C-terminal cleavage and TAP transport efficiency. It has updated the version from 1.0 to 1.2 to improve the accuracy of MHC class I peptide-binding affinity and proteasomal cleavage prediction. NetMHC server 3.0[85] (http://www.cbs.dtu.dk/ services/NetMHC) is based on ANN and weight matrices. It has been trained on data from 55 MHC peptides (43 human and 12 non-human) and position-specific scoring matrices for a further 67 HLA alleles.

MHC class I molecule motifs are well defined but the prediction of MHC class II binding peptides is found to be difficult for a number of reasons, including variable length of reported binding peptides, undetermined core region for each peptide and number of amino acids as primary anchor. Brusic *et al.*[86] developed PERUN, a hybrid method for the prediction of MHC class II binding peptide. It uses available experimental data and expert knowledge of binding motifs, evolutionary algorithms and ANN. They used PlaNet package version 5.6[87] to design and train a three-layered fully connected feed-forward ANN.

Prediction using other machine learning methodologies

Nanni[88] demonstrated the use of SVM and SV (Support Vector) data description to predict T-cell epitopes. In the case of TAPPred[89] (http://www.imtech.res.in/raghav/tappred/), Bhasin and Raghava analysed nine features of amino acids to find the correlation between binding affinity and physiochemical properties. They developed an SVM-based method to predict the TAP binding affinity of peptides, and found cascade SVM to be more reliable. Cascade SVM has two layers of SVMs and its performance is better than the other available algorithms.

Computational techniques are found to be easier than experimental analysis for determining cleavage specificities of proteasomes. It is experimentally established that the immunoproteasome is involved in the generation of the MHC class I ligand. For this purpose, Pcleavage[90] (http://www.imtech.res.in/raghava/pcleavage/) has been developed to predict both kinds of cleavage sites in antigenic proteins. It uses SVM,[91] Parallel Exemplar based Learning[92] and Waikato Environment for Knowledge Analysis.[93]

Ant colony search systems have proved useful for solving combinatorial optimization problems and can be applied to the identification of a multiple alignment of a set of peptides. Basically, they[94] attempt to find an optimal alignment for a given set of peptides based on the search strategy.

Structure-based prediction

Peptide–MHC binding data are necessary to find T-cell epitopes. Current methods are mostly based on peptide binding affinity to MHC for predicting T-cell epitope. The three-dimensional QSAR technology CoMSIA has been applied to the problem of peptide–MHC binding.[95] It uses the interaction potential around aligned sets of three-dimensional peptide structures to describe binding. TEPITOPE[96] by Bian and Hammer is used to predict promiscuous and allele-specific HLA II restricted T-cell epitopes *in silico*. TEPITOPE's user interface has display and comparison of pocket profiles, and finds similar HLA II differing in their binding capacity for a given peptide sequence. Kangueane and Sakharkar[97] implemented a web server T-cell epitope designer for MHC peptide which uses a definition of virtual binding pockets to position specific peptide residue anchors and estimation of peptide residue virtual binding pocket compatibility.

Zhao *et al.*[98] described a novel predictive model using information from 29 human MHCp crystal structures. The overall binding between peptide and MHC provides a cumulative measure of the physical and chemical compatibility between each residue in the peptide and the residue forming the virtual pockets. ElliPro[99] (http://tools.immuneepitope.org/tools/ElliPro) is a web tool that implements a modified version of the Thorton method, residue clustering algorithm, the Modeller program and the Jmol viewer. It predicts and visualizes the antibody epitope in protein sequence and structure. It implements three algorithms for the approximation of the protein shape as an ellipsoid, calculation of the residue protrusion index and clustering of neighbouring residue based on their protrusion index values.

It is generally accepted that only peptides that bind to MHC with an affinity above a threshold value (typically 500 nm), function as T-cell epitopes. Guan *et al.*[100] in the Edward Jenner Institute for Vaccine Research, UK, introduced MHCPred (http://www.darrenflower.info/mhcpred/). It is a Perl implementation of two-dimensional QSAR application to peptide–MHC prediction and covers both class I and class II MHC allele peptide specificity models. Peptides that can bind to MHC on the tumour cell surface have potential to initiate a host immune response against the tumour. Schiewe and Haworth[101] developed an algorithm PeSSI (peptide–MHC prediction of structure through solvated interfaces) for flexible structure prediction of peptide binding to the MHC molecule. They used CT antigens (Cancer Testis), KU-CT-1, that have the potential to bind HLA-A2.

Jojic *et al.*[102] developed an improved structure-based model which used known three-dimensional structures of a small number of MHC–peptide complexes, the MHC class I sequence, known binding energies for MHC–peptide complexes, and a larger binary dataset with information about strong binders and non-binders. They used adaptive double threading, where the parameters of the threading model are learnable, and both MHC and peptide sequences can be threaded onto the structure of other alleles. Furman *et al.*[103] used an approach that can be applied to a wide range of MHC class I alleles. In this algorithm, peptide candidates are threaded, and their binding compatibility is evaluated by statistical pairwise potentials. They used the pairwise potential table of Miyazawa and Jernigan.[104]

Immunodominant peptides are being used for rational design of peptide vaccines focusing on T-cell immunity. Altuvia and Margalit[105] focused on antigenic peptides recognized by cytotoxic T cells. They applied the threading approach to screen a library of peptide sequences and identified those that optimally fitted within the MHC groove. Propred[106] (http://www.imtech.res.in/ raghava/proped) is a graphical web tool for predicting MHC class II binding regions in antigenic protein sequences. They extracted the matrices for 51 HLA-DR alleles from a pocket profile database developed by Sturniolo *et al.*[107] The EpiToolkit[108] (http://www.epitoolkit.org) web server includes several prediction methods for MHC class I and class II ligands, and minor histocompatibility antigens. It can also investigate the effect of mutation on T-cell epitopes.

Allergy prediction

Food derived from biotechnology and genetic engineering contains some foreign proteins, which can be allergic to many human beings. Because of this, food safety is an important issue. Evaluation of the potential allergenicity of food derived from biotechnology and genetic engineering is a current food safety assessment. Allergen sequence databases are essential tools for safety assessments of bioengineered foods. They can analyse the structural and physiochemical properties of food allergen proteins. They focus on molecular information such as protein sequences, structures and biomedical information.

Allergy occurs by both extrinsic and intrinsic factors. A type I hypersensitive reaction is induced by certain allergens that elicit IgE antibodies.[2] Use of genetically modified food and therapeutics makes allergenic protein prediction necessary. According to the proposed guidelines of World Health Organization (WHO) and Food and Agriculture Organization (FAO) in 2001, a protein is considered an allergen when it has at least six contiguous amino acids the same or a window of 80 amino acids when compared with known allergens. It has already been established that allergens do not share common structural characteristics. Hence, allergen databases are being used as reference for finding the sequence similarity in allergenicity evaluation.[109] It is said that a protein is considered an allergen if it has a region or peptides identical to a known IgE epitope.

The allergen prediction method proposed by Kong *et al.*[110] is based on the determination of a combination of two allergen motifs in a given protein sequence. They took 575 proteins for allergen dataset and 700 sequences for a non-allergen test set from the given reference.[111] They developed a database that has all possible combinations of two motifs from the set of allergenic motifs by using a motif length of 35 amino acids and motif number of 500. Zorzet *et al.*[112] introduced a computational approach for classifying the amino acid sequences in allergens and non-allergens. They identified 91 pre-processed food allergens from various specialized public repositories of food allergy and the SWALL database (SWISSPROT and TrEMBL).

Saha and Raghava[113] created AlgPred (http://www.imtech.res.in/raghava/algpred) using SVM and a similarity-based approach for analysis, and scanned all 183 IgE epitopes against all proteins of the dataset. The server allows use of a hybrid option to predict allergens using a combined approach (SVMc, IgE epitope, ARPs BLAST and MAST).

Stadler and Stadler[109] used the MEME motif discovery tool to identify the most relevant motif present in an allergen sequence. If the query finds an allergen motif or scores better than an E-value of 10^{-8} in the pairwise sequence alignment step, it is considered as the allergenic sequence. Then, these are compared with the FAO/WHO guidelines by performing allergenicity prediction for the sequence in SWISSPROT and a synthetic test database. Allermatch[114] (http://www.allermatch.org) is a webtool that uses a sliding window approach to predict potential allergenicity of proteins. It is done according to the current recommendations of the FAO/WHO Expert Consultation,[115] as outlined in Codex alimentarius.[116] But this method generates false-positive and false-negative hits so it is advised by the FAO/WHO that the outcomes should be combined with other allergenicity assessment methods.

The APPEL[117] (Allergen Protein Prediction E-Lab) tool uses SVM to identify novel allergen proteins. This tool correctly classified 93% of 229 allergens and 99·9% of 6717 non-allergens. It is based on a statistical method and has the potential to discover novel allergen proteins. The EVALLER[118] web server (http://bioinformatics.bmc.uu.se/evaller.html) uses a filtered length-adjusted allergen peptides (DFLAP) method[119] (via ulfh@slv.se) to identify the potential allergen proteins. DFLAP extracts variable length allergen sequence fragments and employs SVM. An uncertainty score has shown that the EVALLER is much more confident in identifying the 'presumably an allergen' category than that of non-allergens.

The EVALLER and APPEL servers assigned all calmodulins or calmodulin-like proteins as presumably non-allergens.[118] But a conventional alignment approach (e.g. 35% similarity over 80 amino acid segments) gives preference to finding sequence similarity between input proteins and known allergens and put the above-mentioned proteins the in allergen category. These proteins are presumably non-allergenic homologues to the polcalcin family (members being potential allergens involved in pollen–pollen cross-sensitization). Tools based on structural and physical characteristics are useful to identify potential cross-reacting proteins that may escape detection through the sequence similarity method alone.

Applications of immunoinformatics

In this section, we focus on applications of immunoinformatics. It includes *in silico* vaccine design and immune system modelling.

In silico vaccination

It is easy to apply new approaches for vaccine design, as genome sequencing, comparative proteomics and immunoinformatics tools are well developed. 'Reverse vaccinology', a new concept, analyses the entire genome to identify potentially antigenic extracellular proteins and so helps to save time and money. It was pioneered for *Neisseria meningitides*, which is responsible for sepsis and meningococcal meningitides. The vaccine type is conjugate and is based on capsular polysaccharide. These vaccines are available for pathogenic *N. meningitides* A, C, Y and W135.[120]

Microarray technique for vaccine design

Through microarray technology, it is easy to screen genes of various pathogens in different growth states and conditions for vaccine design.[121] It reduces the number of genes useful for vaccine in a given genome. Signal peptides derived from genomic sequences, structural motifs and immunogenicity are important for vaccine development.

Epitope-driven approaches for vaccine design

These are comparatively more useful as they have no lethal effect like the whole protein vaccines. It may induce an immune response against immunodominant epitopes.[122] This kind of vaccine has a single start codon with an epitope which can be inserted consecutively in the construct.[123] The prediction of promiscuous binding ligands is considered to be a prerequisite for most subunit vaccine design strategies.[124]

Peptide-based vaccine design

Small peptides derived from epitopes are used as peptide-based vaccines. These peptides are recognized by MHC class I and therefore boost the immune response. Florea *et al.*[125] described three novel classes of methods to predict MHC binding peptides, and a voting scheme to integrate them for improved results. The first method is based on quadratic programming applied to quantitative and qualitative data. The second method uses linear programming and the third one considers sequence profiles obtained by clustering known epitopes to score candidate peptides. This method is found to be better than other sequence-based methods for finding the MHC binders.

Alignment-free approach for vaccine design

Earlier approaches for the identification of antigens were dependent on sequence alignment, which had several drawbacks. Some proteins have similar structure and biological properties, but they may lack sequence similarity. To get rid of these limitations, a new alignment-free approach for antigen prediction has been proposed for which Doytchinova and Flower[126] used three datasets, one each for bacteria, viruses and tumours. The models were validated using leave-one-out cross-validation (LOO-CV) on the whole sets and by external validation using test sets. These models were implemented in a server called VaxiJen (http://www.darrenflower.info/VaxiJen/).

DNA vaccines

It has already been found that DNA vaccines can produce both cell-mediated and humoral immune responses, and are very useful in defending intracellular pathogens. DyNAVacs[127] (http://miracle.igib.res.in/dynavac/) incorporates different modules like codon optimization for heterologous expression of genes in bacteria, yeast and plants, mapping restriction enzyme sites, primer design, Kozak sequence insertion, custom sequence insertion and design of genes for gene therapy.

The software NERVE[128] (http://www.bio.unipd.it/molbinfo) helps in designing subunit vaccines against bacterial pathogens. It combines automation with an exhaustive treatment of vaccine candidate selection tasks by implementing and integrating six different kinds of analyses. Xiang *et al.*[129] developed a web-based database system, VIOLIN (Vaccine Investigation and Online Information Network) (http://www.violinet.org), which curates, stores and analyses published vaccine data. It contains four integrated literature mining and search programs: Litsearch, Vaxpresso, Vaxmesh and Vaxlert. They have developed a web-based vaccine design system called Vaxign,[130] which predicts possible vaccine targets. Major predicted features include subcellular location of a protein, transmembrane domain, adhesion probability, sequence conservation among genomes, sequence similarity to host (human or mouse) proteome, and epitope binding to MHC class I and class II.

Immune system modelling

Immune system modelling provides an integrated view of the immune system in both qualitative and quantitative terms. These models can test and find out the antigen–antibody interactions and immune responses for a particular antigen, in case of drug administration or testing of a vaccine candidate. This helps in reducing time and cost. Peters *et al.*[33] developed a hepatitis C virus infection model that could predict the results of tumour necrosis factor-α acting by blocking *de novo* infection, blocking viral replication or effecting virion clearance. A model can calculate the likelihood of HIV developing a drug-resistant mutation, if provided with certain replication and mutation rates. Using the visual modelling application described by Gong and Cai,[131] one can understand the adaptive immune system effectively. The hierarchical immune system consists of an inherent immune tier, an adaptive immune tier and an immune cell tier. It is designed and visualized with the Java Applet technique for simulation. For further simulation purpose, the learning of the antibody is implemented through the evolutionary mechanism of the immune algorithm. ImmunoGrid (http://www.immunogrid.org) and Virolab (http://www.virolab.org:080/virolab) projects are working to simulate immune systems. ImmunoGrid tries to simulate immune processes by combining experiments and computational studies while Virolab is attempting to develop a virtual laboratory for infectious diseases by examining the genetic causes of human illnesses.[121] SIMISYS 0.3[132] is another example of a software that models and simulates the innate and adaptive components of the immune system, based on computational framework of cellular automata. It simulates healthy and disease conditions by interpreting interactions among the cells including, macrophages, dendritic cells, B cells, T helper cells and pathogenic bacteria.

Exclusive computational approaches like mathematical modelling generate enormous amounts of data, but there should be a balance between virtual and real experimental data. Computationally generated data need to be formally tested and translated into real knowledge. The post-genomic era needs to exchange data from wet laboratory to simulation and vice versa.[133] The model should be accurate, easy to use and understandable to both model designers and biologists, who can verify their hypothesis through *in silico* experiments.

Conclusions and discussions

This review considers useful online immunological databases, tools and webservers. It is described how immunoinformatics is useful in reducing the time and cost involved in the traditional study of immunology. Immunoinformatics may be placed at the junction point between experimental and computational approaches. It complements wet laboratory immunology.

Most of the existing methods tend to predict epitopes with high affinity to MHC molecules. These methods are indirect as they predict MHC binders instead of T-cell epitopes, as opposed to the earlier methods. It is hypothesized that the T cell recognizes a peptide of amphipathic nature. The hydrophobic terminal of the antigenic peptide reacts with MHC while the hydrophilic end interacts with the TR. Earlier approaches used this phenomenon. Methods based on predicting structural binding motifs need structural data generated by molecular biology. This approach scans epitopic sequences to find MHC binders. However, these approaches become useless if motifs are not present. They need the three-dimensional structure of the MHC–peptide complex, which is again a limitation.

A matrix-driven method needs information about each residue of interacting peptide, and thereby gives better results. Machine-learning techniques are quite good as they can deal with non-linear data. Earlier approaches have some limitations in handling real data (non-linear data). SVM (a statistical learning methodology) is a learning technique that supports continuous and categorical variables. SVM is better than ANN because it attains a global minimum and is capable of working with fewer training patterns.[134] Hence both sequence characteristics and computational techniques should be integrated to acquire higher prediction accuracy. Recently, the prediction of promiscuous peptides (capable of binding to a wide array of MHC molecules) is being given much emphasis. Screening of large-scale pathogens and mapping of T-cell epitopes allow identification of the prime target of epitope-based T-cell vaccine designs.

'Reverse vaccinology' is a revolution in immunology because it uses the whole spectrum of antigens. This helps in using pools of vaccine candidates that otherwise would be missed (because of poor or no *in vitro* experimental information or problems in culturing the specific pathogen).[134] It makes the available pools of vaccine candidates easier to use when designing therapeutic vaccines. As of now, different groups are applying reverse vaccinology approaches that show promising pre-clinical results.

Immunoinformatics models are being used that are analogous to and that simulate the real behaviour of immune system processes. These models help in understanding the kinetics of cells during immune responses. They make understanding the biological pathways and underlying mechanisms easier. The models are engineered in such a way that they can be studied and interpreted easily, and can be rebuilt if new experimental data are introduced. These mathematical models remove the uncertainty of systems; as they are found to be close to wet laboratory experiments this leads to designing the path for refinement and modelling new experiments. Computational modelling of the immune system provides scientific solutions to several problems but it should not be forgotten that they rely on assumptions only, so they cannot be directly compared with real biological data. They can be improved by the availability of more data, significant parameters, or by modifying the underlying equations. These changes can better mimic the biological interactions in an organism. Currently, models are designed to simulate the biological data only over a fixed time period.[135] There are no data for extended time spans available to validate the models. This limits the accuracy of the results. An ability of these models to show the system's changes over an extended time period for immune response in case of antigen attack or drug administration would reduce the necessity for experimental research.

Exploration of the immune response to a specific drug can be a future research area in the modelling field. Drug response to a host's immune system can be better studied through computational models. The effect of drug administration can be added to model the immune system to find the drug efficacy.[135]

Moreover, the field of immune system modelling provides ideas about the dose composition, drug dosage duration, age of the patient and other parameters. It can give new suggestions for the study of immune system function and drug function to treat certain diseases. These modelling capabilities may lead to the invention of drugs that can treat a disease in a more effective way and without any side-effects. Diseases that are characterized by complex interactions between the host cellular immune system and evolving pathogens such as HIV infection can be investigated by such models.

Disclosures

The authors have no financial disclosures.

References

1 Kimbrell DA, Beutler B. The evolution and genetics of innate immunity. *Nat Rev Genet* 2001; **2**:256–67.

2 Thomas K, Goldsby J, Osborne RA, Barbara A, Kuby J. *Kuby Immunology*, 6th edn. New York: WH Freeman and Co, 2006.

3 Korber B, LaBute M, Yusim K. Immunoinformatics comes of age. *PLoS Comput Biol* 2006; **2**:0484–92.

4 Gardy JL, Lynn DJ, Brinkman FSL, Hancock REW. Enabling a systems biology approach to immunology: focus on innate immunity. *Trends Immunol* 2009; **30**:249–62.

5 Davies MN, Flower DR. Harnessing bioinformatics to discover new vaccine. *Drug Discov Today* 2007; **12**:389–95.

6 Ortutay C, Vihinen M. Immunome Knowledge base (IKB): an integrated service for immunome research. *BMC Immunol* 2009; **10**. doi:10.1186/1471-2172-10-3.

7 Sette A, Fleri W, Peters B, Sathiamurthy M, Bui HH. A roadmap for the immunomics of category A-C pathogens. *Immunity* 2005; **22**:155–61.

8 Groot ASDe. Immunomics: discovering new targets for vaccine and therapeutics. *Drug Discov Today* 2006; **11**:203–9.

9 Grainger DJ. Immunomics: principles and practice. *IRTL* 2004; **2**:1–6.

10 Yates A, Chan CCW, Callard RE, George AJT, Stark J. An approach to modelling in immunology. *Brief Bioinform* 2001; **2**:245–57.

11 Kaplan No, Everse J, Dixon Je, Stolzenbach Fe, Lee Cy, Lee Clt, Taylor Ss, Mosbach K. Purification and separation of pyridine nucleotide-linked dehydrogenases by affinity chromatography techniques. *Proc Natl Acad Sci USA* 1974; **71**:3450–4.

12 Davey HM. Flow cytometric techniques for the detection of microorganisms. *Methods Cell Sci* 2004; **24**:91–7.

13 Durkin MM, Patricia A, Connolly PA, Wheat LJ. Comparison of radioimmunoassay and enzyme-linked immunoassay methods for detection of *Histoplasma capsulatum* var. *capsulatum* antigen. *J Clin Microbiol* 1997; **35**:2252–5.

14 Ma H, Shieh KJ, Lee SL. Study of ELISA technique. *Nat Sci* 2006; **4**:36–7.

15 Nishimaki T, Sagawa K, Motogi S, Saito K, Morito T, Yoshida H, Kasukawa R. A competitive inhibition test of enzyme immunoassay for the anti-nRNP antibody. *J Immunol Methods* 1987; **100**:157–60.

16 Levine MA, Thornton P, Forman SJ *et al.* Positive Coombs test in Hodgkin's disease: significance and implications. *Blood* 1980; **55**:607–11.

17 Wanga B, Huaa RH, Tiana Z-J, Chena N-S, Zhaoa F-R, Liua T-Q, Wanga Y-F, Tong G-Z. Identification of a virus-specific and conserved B-cell epitope on NS1 protein of Japanese encephalitis virus. *Virus Res* 2009; **141**:90–5.

18 Admon A, Barnea E, Ziv T. Tumor antigens and proteomics from the point of view of the major histocompatibility complex peptides. *Mol Cell Proteomics* 2003; **2**:388–98.

19 Boon T, Coulie PG, den Eynde BV. Tumor antigens recognized by T cells. *Immunol Today* 1997; **18**:267–8.

20 Gubler DJ. Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev* 1998; **11**:480–96.

21 Amin N, Aguilar A, Chamac ho F *et al.* Identification of dengue-specific B-cell epitopes by phage-display random peptide library. *Malaysian J Med Sci* 2009; **16**:4–14.

22 Wang Y. Immunostaining with dissociable antibody microarrays. *Proteomics* 2004; **4**:20–6.

23 Magdalena J, Odling J, Qiang PH, Martenn S, Joakin L, Uhlen M, Hammarstrom L, Nilsson P. Serum microarrays for large scale screening of protein levels. *Mol Cell Proteomics* 2005; **4**:1942–7.

24 Sahin U, Tureci O, Pfreundschuh M. Serological identification of human tumor antigens. *Curr Opin Immunol* 1997; **9**:709–16.

25 Oelke M, Maus MV, Didiano D, June CH, Mackensen A, Schneck JP. *Ex vivo* induction and expansion of antigen-specific cytotoxic T cells by HLA-Ig coated artificial antigen-presenting cells. *Nat Med* 2003; **9**:619–24.

26 Groot DeAS, Shai H, Aubin CS, Mcmurry J, Martin W. Immuno-informatics: mining genomes for vaccine components. *Immunol Cell Biol* 2002; **80**:225–69.

27 Quintana FJ, Hagedorn PH, Gad E, Yifat M, Eutan D, Cohen IR. Functional immunomics: microarray analysis of IgG autoantibody repertoires predicts the future response of mice to induced diabetes. *Proc Natl Acad Sci USA* 2004; **101**:14615–21.

28 Sampson HA. Food allergy – accurately identifying clinical reactivity. *Allergy* 2005; **60**:19–24.

29 Vegvar de HEN, Robinson WH. Microarray profiling of antiviral antibodies for the development of diagnostics, vaccines, and therapeutics. *J Clin Immunol* 2004; **111**:196–201.

30 Henry ENdeV, RamaRao A, Lawrence S, Paul JU, Harriet LR, Robinson WH. Microarray profiling of antibody responses against simian-human immunodeficiency virus: postchallenge convergence of reactivities independent of host histocompatibility type and vaccine regimen. *J Virol* 2003; **77**:11125–38.

31 Nahtman T, Jernberg A, Mahdavifar S, Zerweck J, Schutkowski M, Maeurer M, Reilly M. Validation of peptide epitope microarray experiments and extraction of quality data. *J Immunol Methods* 2007; **328**:1–13.

32 Braga-Neto UM, Marques ETA. From functional genomics to functional immunomics: new challenges, old problems, big rewards. *PLoS Comput Biol* 2006; **2**:651–62.

33 Peters B, Sidney J, Bourne P *et al.* The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol* 2005; **3**:1361–70.

34 Lynn DJ, Winsor GL, Chan C *et al.* InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol Syst Biol* 2008; **4**:1–11.

35 Barsky S, Gardy JL, Hancock R, Munzer T. Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. *Bioinformatics* 2007; **23**:1040–2.

36 Shanon P, Markiel A, Ozier O *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**:2498–504.

37 Saha S, Bhasin M, Raghava GPS. Bcipep: a database of B-cell epitopes. *BMC Genomics* 2005; **6**. doi:10.1186/1471-2164-6-79.

38 Huang J, Honda W. CED: a conformational epitope. *BMC Immunol* 2006; **7**:7.

39 Schlessinger A, Ofran Y, Yachdav G, Rost B. Epitome: database of structure-inferred antigenic epitopes. *Nucleic Acids Res* 2006; **34**:D777–80.

40 Blythe MJ, Doytchinova IA, Darren R. JenPep: a database of quantitative functional peptide data for immunology. *Bioinformatics* 2002; **18**:434–9.

41 Rammensee HG, Bachmann J, Emmerich NPN, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 1999; **50**:213–9.

42 Feldhahn M, Donnes P, Thiel P, Kohlbacher O. FRED – a framework for T-cell epitope detection. *Bioinformatics* 2009; **25**:2758–9.

43 Lefranc M-P, Giudicelli V, Ginestoux C *et al.* IMGT®, the international ImMunoGeneTics information system®. *Nucleic Acids Res* 2009; **37**:D1006–12.

44 Lord P, Shah N, Sansone SA, Stephens S, Soldatova L (eds). The OBI Consortium. Modeling biomedical experimental processes with OBI, In: *Proceedings of the 12th Annual Bio-Ontologies 35 Meeting. International Society for Computational Biology*. Sweden: Stockholm 2009;41–4.

45 Mari A, Scalab E, Palazzob P, Ridolfib S, Zennarob D, Carabella G. Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. *Cell Immunol* 2006; **244**:97–100.

46 Pomes A. Relevant B cell epitopes in allergic disease. *Int Arch Allergy Immunol* 2010; **152**:1–11.

47 Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TAE, Thomas W. Allergen nomenclature. *Bull World Health Organ* 1994; **72**:796–806.

48 Kim C, Kwon S, Lee G, Lee H, Choi J, Kim Y, Hahn J. A database for allergenic proteins and tools for allergenicity prediction. *Bioinformation* 2009; **3**:344–5.

49 Ivanciuc O, Schein CH, Braun W. SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res* 2003; **31**:359–62.

50 Ortutay C, Siermala M, Vihinen M. ImmTree: database of evolutionary relationships of genes and proteins in the human immune system. *Immunome Res* 2007; **3**: doi: 10.1186/1745-75803-4.

51 Ortutay C, Vihinen M. Immunome: a reference set of genes and proteins for systems biology of the human immune system. *Cell Immunol* 2006; **244**:87–9.

52 Rannikko K, Ortutay C, Vihinen M. Immunity genes and their orthologs: a multi-species database. *Int Immunol* 2007; **19**:1361–70.

53 Greenbaum JA, Andersen PH, Blythe M *et al.* Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J Mol Recognit* 2007; **20**:75–82.

54 Tong JC, Ren EC. Immunoinformatics: current trends and future directions. *Drug Discov Today* 2009; **14**:684–9.

55 Pellequer J, Westhof E, Regenmortel MV. Predicting the location of structure of continuous epitopes in proteins from their primary structure. *Methods Enzymol* 1991; **203**:176–201.

56 Parker J, Guo D, Hodges R. 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* 1986; **25**:5425–32.

57 Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol* 1978; **47**:45–148.

58 Levitt M. Conformational preferences of amino acids in globular proteins. *Biochemistry* 1978; **17**:4277–85.

59 Emini E, Hughes J, Perlow D, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus specific synthetic peptide. *J Virol* 1985; **55**:836–9.

60 El-Manzalawy Y, Dobbs D, Honavar V. Predicting protective linear B-cell epitopes using evolutionary information. In: *Proceedings of the 2008 IEEE International Conference on Bioinformatics and Biomedicine*, Washington: IEEE Computer Society 2008:289–92.

61 Odorico M, Pellequer JL. BEPITOPE: predicting the location of continuous epitopes and patterns in protein. *J Mol Recognit* 2003; **16**:20–2.

62 Pellequer JL, Westhof E. PREDITOP: a program for antigenicity predictions. *J Mol Graph* 1993; **11**:204–10.

63 Saha S, Raghava GPS. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In: Nicosia G, Cutello V, Bentley PJ, Timis J eds. *Artificial Immune Systems.* Berlin/Heidelberg: ICARIS Springer, LNCS 2004; 3239:197–204.

64 Ghate AD, Bhagwat BU, Bhosle SG, Gadepalli SM, Kulkarni-Kale UD. Characterization of antibody-binding sites on proteins: development of a knowledge base and its applications in improving epitope prediction. *Protein Pept Lett* 2007; **14**:531–5.

65 Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 2006; **65**:40–8.

66 Sweredoski MJ, Baldi P. COBEpro: a novel system for predicting continuous B-cell epitopes. *Protein Eng Des Sel* 2009; **22**:113–20.

67 Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B cell epitopes. *Immunome Res* 2006; doi:10.1186/1745-7580-2-2.

68 Toseland CP, Clayton DJ, McSparron H *et al.* AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res* 2005; **1**. doi:10.1186/1745-7580-1-4.

69 Evans MC. Recent advances in immunoinformatics: application of *in silico* tools to drug development. *Curr Opin Drug Discov Devel* 2008; **11**:233–41.

70 Anderson Ph, Nielsen M, Lund O. Prediction of residues in discontinuous B cell epitopes using protein 3D structures. *Protein Sci* 2006; **15**:2558–67.

71 Bublil EM, Mayrose NTFI, Penn O, Berman AR. Stepwise prediction of conformational discontinuous B-cell epitopes using the mapitope algorithm. *Proteins* 2007; **68**:294–304.

72 Sollner J, Grohmann R, Rapberger R, Perco P, Lukas A, Mayer B. Analysis and prediction of protective continuous B cell epitopes on pathogen proteins. *Immunome Res* 2008; **4**. doi:10.1186/1745-7580-4-1.

73 Kale KU, Bhosle S, Kolaskar AS. CEP: a conformational epitope prediction server. *Nucleic Acids Res* 2005; **33**:W168–71.

74 Mayrose I, Penn O, Erez E *et al.* Pepitope: epitope mapping from affinity-selected peptides. *Bioinformatics* 2007; **23**:3244–6.

75 Moreau V, Granier C, Villard S, Laune D, Molina F. Discontinuous epitope prediction based on mimotope analysis. *Bioinformatics* 2006; **22**:1088–95.

76 Huang J, Gutteridge A, Honda W, Kanehisa M. MIMOX: a web tool for phage display based epitope mapping. *BMC Bioinformatics* 2006; **7**. doi: 10.1186/1471-2105-7-451.

77 Mayrose I, Shlomi T, Rubinstein ND, Gershoni JM, Ruppin E, Sharan R, Pupko T. Epitope mapping using combinatorial phage-display libraries: a graph-based algorithm. *Nucleic Acids Res* 2007; **35**:69–78.

78 Schreiber A, Humbert M, Benz A, Dietrich U. 3D-Epitope-Explorer (3DEX): localization of conformational epitopes within three-dimensional structures of proteins. *J Comput Chem* 2005; **26**:879–87.

79 Huang L, Dai Y. Direct prediction of T-cell epitopes using support vector machines with novel Sequence encoding schemes. *J Bioinform Comput Biol* 2006; **4**:93–107.

80 Bhasin M, Raghava GPS. Prediction of promiscuous and high-affinity mutated MHC binders. *Hybrid Hybridomics* 2003; **22**:229–34.

81 Zhang GL, Petrovsky N, Kwoh CK, August JT, Brusic V. Pred^{TAP} : a system for prediction of peptide binding to the human transporter associated with antigen processing. *Immunome Res* 2006; **2**. doi: 10.1186/1745-7580-2-3.

82 Neilsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, Brunak S, Lund O. Reliable prediction of T-cell epitopes using networks with novel sequence representations. *Protein Sci* 2003; **12**:1007–17.

83 Buus S, Stryhn A, Winther K, Kirkby N, Pedersen LO. Receptor–ligand interactions measured by an improved spun column chromatography technique. A high efficiency and high throughput size separation method. *Biochim Biophys Acta* 1995; **1243**:453–60.

84 Larsen MV, Lundegaard C, Lamberth K, Buss S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 2007; **8**. doi: 10.1186/1471-2105-8-424.

85 Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res* 2008; **36**:W509–12.

86 Brusic V, Rudy G, Honeyman M, Hammer J, Harrison L. Prediction of MHC class II-binding peptides using an evolutionary and artificial neural network. *Bioinformatics* 1998; **14**:121–30.

87 Miyata J. A User's Guide to PlaNet Version 5.6. Boulder: Computer Science Department, University of Colorado. 1991.

88 Nanni L. Machine learning algorithms for T-cell epitopes prediction. *Neurocomputing* 2006; **69**:866–8.

89 Bhasin M, Raghava GPS. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci* 2004; **13**:596–607.

90 Bhasin M, Raghava GPS. Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. *Nucleic Acids Res* 2005; **33**:W202–7.

91 Joachims T. Marking large-scale support vector machine learning practical. In: Scholkopf B, Burges CJC, Smola AJ, eds. *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA, USA: MIT Press, 1999:169–84.

92 Cost S, Salzberg S. A weighted nearest neighbor algorithm for learning with symbolic features. *Mach Learn* 1993; **10**:57–78.

93 Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, 2nd edn. San Francisco: Morgan Kaufman, 1999.

94 Dorigo M, Maniezzo MV, Cambini AC. Ant system: optimization by a colony of cooperating agents. *IEEE Trans Syst Man Cybern Part B* 1996; **26**:29–41.

95 Flower DR. Towards *in silico* prediction of immunogenic epitopes. *Trends Immunol* 2003; **24**:667–74.

96 Bian H, Hammer H. Discovery of promiscuous HLA restricted T cell epitope with TEPITOPE. *Methods* 2004; **34**:468–75.

97 Kangueane P, Sakharkar MK. T epitope designer: HLA peptide binding prediction server. *Bioinformation* 2005; **1**:21–4.

98 Zhao B, Mathura VS, Ganapathy R, Moochhala S, Sakharkar MK, Kangneane P. A novel MHCp binding prediction model. *Hum Immunol* 2003; **64**:1123–43.

99 Ponomarenko JV, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, Peters B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* 2008; **9**. doi: 10.1186/1471-2105-9-514.

100 Guan P, Doytchinova IA, Zygouri C, Flower DR. MHCPred: a server for quantitative prediction of peptide-MHC binding. *Nucleic Acids Res* 2003; **31**:3621–4.

101 Schiewe AJ, Haworth IS. Structure based prediction of MHC–peptide association: algorithm comparison and approach to cancer vaccine design. *J Mol Graph Model* 2007; **26**:667–75.

102 Jojic N, Gomez MR, Heckerman D, Kadie C, Furman OS. Learning MHC-I peptide binding. *Bioinformatics* 2006; **22**:e227–35.

103 Furman OS, Altuvia Y, Sette A, Margalit H. Structure-based prediction of binding peptides to MHC class I molecules: application to a broad range of MHC alleles. *Protein Sci* 2000; **9**:1838–46.

104 Miyazawa S, Jernigan RL. Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol* 1996; **256**:623–44.

105 Altuvia Y, Margalit H. A structure-based approach for prediction of MHC-binding peptides. *Methods* 2004; **34**:454–9.

106 Singh H, Raghava GPS. Propred: prediction of HLA-DR binding sites. *Trends Immunol* 2001; **17**:1236–7.

107 Sturniolo T, Bono E, Ding J *et al.* Generation of tissue-specific and promiscuous HLA ligand database using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* 1999; **17**:555–61.

108 Feldhahn M, Thiel P, Schuler MM, Hillen N, Stevanovic S, Rammensee HG, Ohlbacher O. EpiToolKit – a web server for computational immunomics. *Nucleic Acids Res* 2008; **1**:W519–22.

109 Stadler MB, Stadler BM. Allergenicity prediction by protein sequence. *FASEB J* 2003; **17**:1141–3.

110 Kong W, Tan TS, Tham L, Choo KW. Improved prediction of allergenicity by combination of multiple sequence motifs. *In Silico Biol* 2006; **7**:77–86.

111 Bjorklund AK, Atmadja SD, Zorzet A, Hammerling U, Gustafsson MG. Supervised identification of allergen-representative peptides for *in silico* detection of potentially allergenic proteins. *Bioinformatics* 2005; **21**:39–50.

112 Zorzet A, Gustafsson M, Hammerling U. Prediction of food protein allergenicity: a bioinformatic learning systems approach. *In Silico Biol* 2002; **2**:525–34.

113 Saha S, Raghava GPS. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res* 2006; **34**:W202–9.

114 Fiers MWEJ, Kleter GA, Nijland H, Peijnenburg AACM, Peter NJ, Ham RCHJV. Allermatch™, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics* 2004; **5**. doi: 10.1186/1471-2105-5-133.

115 FAO/WHO. Allergenicity of Genetically Modified Foods. 2001; Available at http://www.who.int/foodsafety/publications/biotech/en/ec_jan2001.pdf.

116 FAO/WHO. Codex Principles and Guidelines on Foods Derived from Biotechnology. 2003; Available at ftp://ftp.fao.org/codex/standard/en/CodexTextsBiotechFoods.pdf.

117 Cui J, Han LY, Li H, Ung CY, Tang ZQ, Zheng CJ, Cao ZW, Chen YZ. Computer prediction of allergen proteins from sequence-derived protein structural and physicochemical properties. *Mol Immunol* 2007; **44**:514–20.

118 Barrio AM, Atmadja DS, Nistr A, Gustafsson MG, Hammerling U, Rudloff EB. EVALLER: a web server for *in silico* assessment of potential protein allergenicity. *Nucleic Acids Res* 2007; **35**:694–700.

119 Soeria-Atmadja D, Lundell T, Gustafsson MG, Hammerling U. Computational detection of allergenic proteins attains a new level of accuracy with *in silico* variable-length peptide extraction and machine learning. *Nucleic Acids Res* 2006; **34**:3779–93.

120 Pizza M, Scarlato V, Masignani V *et al.* Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 2000; **287**:1816–20.

121 Groot ASDe, Rappuoli R. Genome derived vaccines. *Expert Rev Vaccines* 2003; **3**:59–76.

122 Gallimore A, Hengartner H, Zinkernagel R. Hierarchies of antigen-specific cytotoxic T cell responses. *Immunol Rev* 1998; **164**:29–36.

123 Morris S, Kelly C, Howard A, X Li, Collins F. The immunogenicity of single and combination DNA vaccines against tuberculosis. *Vaccine* 2000; **18**:2155–63.

124 Zhao B, Sakharkar KR, Lim CS, Kangueane P, Sakharkar MK. MHC–peptide binding prediction for epitope based vaccine design. *Int J Integr Biol* 2007; **1**:127–40.

125 Florea L, Haldorsson B, Kohlbacher O, Schwarty R, Hoffman S, Istrail S. Epitope prediction algorithm for peptide-based vaccine design. In *Proceedings of the IEEE Computer Society Conference on Bioinformatics*, Washington: IEEE Computer Society. 2003:17–26.

126 Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumor antigens and subunit vaccines. *BMC Bioinformatics* 2007; **8**. doi: 10.1186/1471-2105-8-4.

127 Nagarajan H, Gupta R, Agarwal P, Scaria V, Pillai B. DyNAVacS: an integrative tool for optimized DNA vaccine design. *Nucleic Acids Res* 2006; **34**:W264–6.

128 Vivona S, Bernante F, Filippini F. NERVE: new enhanced reverse vaccinology environment. *BMC Biotechnol* 2006; **6**. doi: 10.1186/1472-6750-6-35.

129 Xiang Z, Todd T, Ku KP *et al.* VIOLIN: vaccine investigation and online information network. *Nucleic Acids Res* 2008; **36**:D923–8.

130 Xianga Z, He Y. Vaxign: a web-based vaccine target design program for reverse vaccinology. *Procedia in Vaccinology* 2009; **1**:23–9.

131 Gong T, Cai Z. Visual modeling and simulation of adaptive immune system. In *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, China, 2005; 6:6116–9.

132 Kalita JK, Chandrashekar K, Hans R, Selvam P, Newell MK. Computational modelling and simulation of the immune system. *Int J Bioinform Res Appl* 2006; **2**:63–88.

133 Castiglione F, Liso A. The role of computational models of the immune system in designing vaccination strategies. *Immunopharmacol Immunotoxicol* 2005; **27**:417–32.

134 Vivona S, Gardy JL, Ramachandran S, Brinkman FSL, Raghava GPS, Flower DR, Filippini F. Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. *Trends Biotechnol* 2008; **26**:190–200.

135 Daz P, Gillespie M, Krueger J, Prez J, Radebaugh A, Shearman T, Vo G, Wheatley C. A mathematical model of the immune system's response in obesity-related chronic inflammation. In: *McNair/MAOP Summer Research Symposium, Virginia Tech, Blacksburg VA*. 2008; 2:26–45.