Chapter 3

Immunoinformatics: A Brief Review

Namrata Tomar and Rajat K. De

Abstract

A large volume of data relevant to immunology research has accumulated due to sequencing of genomes of the human and other model organisms. At the same time, huge amounts of clinical and epidemiologic data are being deposited in various scientific literature and clinical records. This accumulation of the information is like a goldmine for researchers looking for mechanisms of immune function and disease pathogenesis. Thus the need to handle this rapidly growing immunological resource has given rise to the field known as immunoinformatics. Immunoinformatics, otherwise known as computational immunology, is the interface between computer science and experimental immunology. It represents the use of computational methods and resources for the understanding of immunological information. It not only helps in dealing with huge amount of data but also plays a great role in defining new hypotheses related to immune responses. This chapter reviews classical immunology, different databases, and prediction tool. Further, it briefly describes applications of immunoinformatics in reverse vaccinology, immune system modeling, and cancer diagnosis and therapy. It also explores the idea of integrating immunoinformatics with systems biology for the development of personalized medicine. All these efforts save time and cost to a great extent.

Key words Systems biology, Immunomics, In silico models, T cells, B cells, Allergy, Reverse vaccinology, Personalized medicine

1 Introduction

The human immune system is very complex and operates at multiple levels, viz., molecules, cells, organs, and organisms. Each individual has a unique immune system and will respond differently to immune challenges. It has a combination of biological structures and processes within an organism to protect it against disease. The earliest literary reference to immunology goes back to 430 b.c., courtesy Thucydides [1]. In 1798, Edward Jenner found some milkmaids immune to smallpox because earlier they contacted 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 a dose of vaccine to a rabid dog-bitten boy who later survived. But Pasteur could not explain its mechanism.

Rajat K. De and Namrata Tomar (eds.), *Immunoinformatics*, Methods in Molecular Biology, vol. 1184,
DOI 10.1007/978-1-4939-1115-8_3,

In 1890, experiments of Emil Von Behring and Shibasaburo Kitasato led to the understanding of the mechanism of immunity. Their experiments described that antibodies present in the serum provided protection against pathogens [1].

According to the traditional dogma of immunology, vertebrates have both innate and adaptive immunology. Innate immune system acts more rapidly and is older and more evolutionarily conserved in comparison with adaptive immune system. It provides the backbone on which adaptive immune system was able to evolve. Innate immune system is less specific and works as a first line of defense [2]. It comprises four types of defensive barriers, viz., anatomic (e.g., skin and mucous membranes), physiologic (e.g., temperature, low pH), phagocytic (e.g., blood monocytes, neutrophils, tissue macrophages), and inflammatory (e.g., serum proteins). Adaptive immune responses in vertebrates are generated within 5 or 6 days after the initial exposure to the pathogen. It is coordinated by a network of highly specialized cells that communicate through cell surface molecular interactions and a complex set of intercellular communication molecules known as cytokines and chemokines. Later exposure to the same pathogen induces a heightened and more specific response because it retains memory [1]. Adaptive immune system has two parts: the cellular immune response of T cells and humoral response of B cells [1, 3]. An antigen has a specific small part, known as epitope that 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 the other of the two glycoproteins on their surface, designated as CD8 or CD4. CD4 T cells function as T helper (Th) cells that recognize peptides displayed by MHC class II molecules. On the other hand, CD8 functions as Tc (cytotoxic T) cells which recognize peptides displayed by MHC class I molecules.

The complexity of the immune system arises from its hierarchical and combinatorial properties. Thus huge amount of data related to immune systems is being generated. Immunologic research needs to deal with this complexity. Immunologists have been using high-throughput experimental techniques for quite a long time, which have generated a vast amount of functional, clinical, and epidemiological data. Therefore the development of new computational approaches to store and analyze these data is needed. This gives rise to the field called immunoinformatics. Immunogenomics, immunoproteomics, epitope prediction, and in silico vaccination are different areas of computational immunological research. Recently, systems biology approaches are being applied to investigate the properties of dynamic behavior of an immune system network [4].

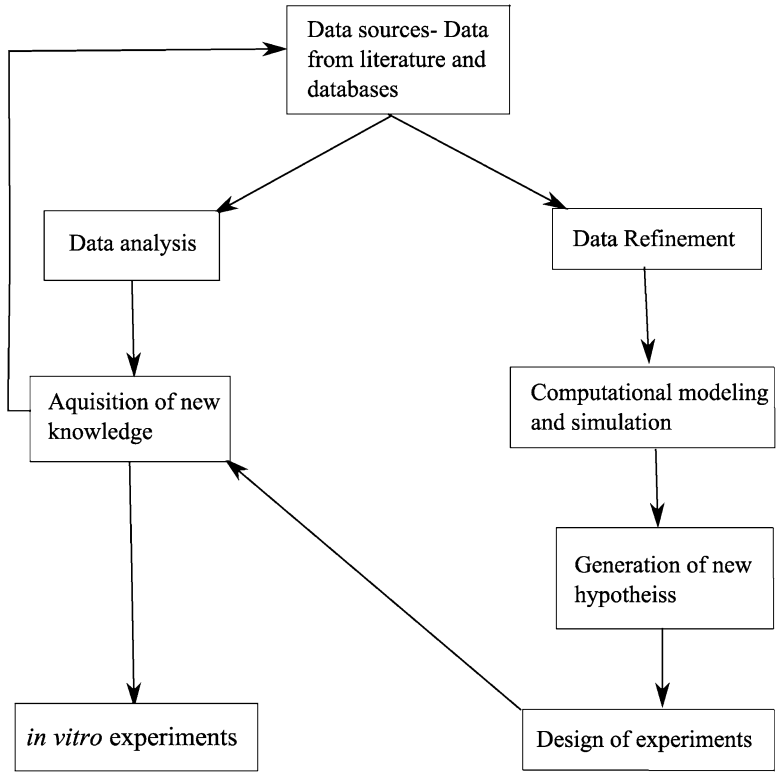

Fig. 1 A possible work flow in immunomics

It includes the study and design of algorithms for mapping potential B and T cell epitopes. These also can lead to exploring the potential binding sites for the development of new vaccines. This methodology is termed as "reverse vaccinology" [5]. It is quite advantageous because conventional methods need to cultivate pathogen and then to extract its antigenic proteins.

All the genes and proteins taking part in immune responses are referred to as "immunome," and it excludes genes and proteins that are expressed in cell types other than in immune cells [6]. All immune reactions due to interaction between host and antigenic peptides are referred to as "immunome reactions," and their study is called as "immunomics" [7]. Like genomics and proteomics, immunomics is a new discipline, which uses high-throughput techniques to understand immune system mechanism [8, 9]. Figure 1 shows work flow in immunomics. This chapter describes various available information regarding classical immunology, different immunomic databases, B and T cell epitope prediction tools and software, and applications of immunoinformatics.

2 Data Sources

Data sources include availability of data from lab experiments through scientific literature, molecular databases, tools and web servers, and clinical records. In this section, we focus on various immune system-related data types and databases. The section starts with some experimental techniques and results.

2.1 Data from Lab Experiments

Immunological experimental and high-throughput molecular biology techniques help in finding the structure and function of immune genes and their products and thereby accumulating a vast amount of experimental data. Experiments involve many immunological techniques to understand the mechanism of an immune system and its responses to various infections, diseases, and drugs, viz., affinity chromatography [10], flow cytometry [11], radioimmunoassay (RIA) [12], enzyme-linked immunosorbent assay (ELISA), [12, 13], competitive inhibition assay [14], and Coombs test [15]. Here, we present some experimental findings that help to identify B and T cell epitopes and to study immune responses.

The ability to identify epitopes in the immune response has important implications in diagnosis of diseases. Thus epitopes for B and T cells need to be identified and mapped. In this context, Wanga et al. [16] mapped B cell epitope present on nonstructural protein (NS1), viz., 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 (GST) tag and then screened by a monoclonal antibody (mAb). They found that the motif of (146) EHARW (150) was the minimal unit of the linear epitope recognized by that mAb. Purification techniques like affinity chromatography are used to purify MHC–peptide from membrane MHC molecules, which can be analyzed by capillary high-pressure liquid chromatography electrospray ionization-tandem mass spectrometry [17]. They can be further used to find new tumor-associated antigens (TAA). One such approach to find TAA is based on transfection of expression library made from cDNA into cells expressing the desired MHC haplotypes [18]. The clones are selected on the basis of their ability to provoke immune response in T cells of the individuals with the same MHC type.

2.2 Exploring the Microarray Technology for Immunomics

"Immunomic microarray" is a microarray technique based on the principle of binding and measurement of target biological specimens to complementary probes. It helps in selecting proteins that cause autoimmunity from genomic sequences [19]. It is being applied to autoimmune disease diagnosis and treatment [20], allergy prediction [21], T and B cell epitope mapping [22], and vaccination [23] to name a few. It includes dissociable antibody

microarray [24], serum microarray [25], and serological analysis of cDNA expression library (SEREX) [26]. An antibody microarray is used to measure concentration of antigen for a specific antibody probe and thereby consists of antibody probes and antigen targets. On the contrast, peptide microarray uses antigen peptides as fixed probes and serum antibodies as targets. The recent technology is peptide–MHC microarray or artificial antigen-presenting chip. In this technique, recombinant peptide–MHC complexes and co-stimulatory molecules are immobilized on a surface, and population of T cells is incubated with the microarray. The T cell spots act as artificial antigen-presenting cells [27] containing a defined MHC-restricted peptides. The advantage of using peptide–MHC is that it can map MHC-restricted T cell epitope.

The immunomic and genomic microarray data have some similarities, yet both of them also differ in several ways; for example, 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 [28, 29]. DNA microarrays measure one response value for each gene per sample; that is, mRNA concentration produced by the gene but a single epitope can generate different response values corresponding to different epitopes in peptide–MHC chips. In case of B cell epitope, it can be recognized by different isotypes of immunoglobulins, so here, one can measure both intensity and quality of antibody response.

2.3 Immunomic Databases

The property of an antigen to bind specifically complementary antibodies is known as the antigen's antigenicity. Likewise, the ability of an antigen to induce an immune response is called its immunogenicity. Immunomic databases include epitope information-related databases, analysis tools, and prediction algorithms, which are crucial for basic immunological studies, diagnosis, and treatment of various diseases and in vaccine research [30]. InnateDB [31] (http://www.innatedb.ca) has been created to understand complete network of pathways and interactions of innate immune system responses. It has ~18,000 annotated molecular interactions of relevance to innate immunity and >1,200 genes, involved in innate immunity according to the recent update till February 16, 2012. It has a newer version, called Cerebral [32], which is a Java plug-in for the cytoscape biomolecular interaction viewer version 2.8.2 [33] 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.

2.4 B Cell Epitope Databases

A brief detail on B cell epitope databases is provided here. Readers can find a detailed description in later chapters. Mapping B cell epitopes plays an important role in vaccine design, immunodiagnostic tests, and antibody production. It has been found that 90 %

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

Databases	Names	URLs
B cell epitopes	CED Bcipep Epiotme IEDB IMGT®	http://www.immunet.cn/ced/log.html http://www.imtech.res.in/raghava/bcipep http://www.rostlab.org/services/epitome/ http://www.immuneepitope.org/ http://www.imgt.org
T cell epitopes	Syfpeithi IEDB IMGT®	http://www.syfpeithi.de http://www.immuneepitope.org/ http://www.imgt.org
Allergen	Database of IUIS SDAP	http://www.allergen.org http://www.fermi.utmb.edu/SDAP/
Information related to molecular evolution of immune system components	ImmTree Immunome database ImmunomeBase Immunome Knowledge Base	http://www.bioinf.uta.fi/ImmTree http://www.bioinf.uta.fi/Immunome/ http://www.bioinf.uta.fi/ImmunomeBase http://www.bioinf.uta.fi/IKB/

of B cell epitopes are conformational or discontinuous; however, they may comprise linear amino acid chain of peptides, which is brought closure in 3D space [34]. Bcipep [35] (http://www.imtech.res.in/raghava/bcipep) gives comprehensive information about experimentally verified B cell epitopes and tools for mapping these epitopes on an antigen sequence. Conformational epitope database (CED) [36] has a collection of B cell epitopes from the literature, conformational epitopes defined by methods, like X-ray diffraction, NMR, scanning mutagenesis, overlapping peptides, and phage display. Epitome [37] (http://www.rostlab.org/services/epitome/) contains all known antigen–antibody complex structures. A semiautomated tool has also been developed which identifies the antigenic interactions within the known antigen–antibody complex structures. They compiled these interactions into Epitome. None of the other databases till now explicitly can locate the complementary determining regions (CDRs) or identify the antigenic residues semiautomatically. Epitome update follows update of SCOP; that is, Epitome is updated twice a year as soon as SCOP gets updated.

The difference between Epitome and CED lies in the source of collection of B cell epitopes. Epitome collects B cell epitopes only from PDB structures and includes CDR information. In contrast, CED takes data from the literature and from abovementioned methods. As their sources are different, one can use the complementary information.

2.5 T Cell Epitope Databases

A brief detail on T cell epitope databases is provided here. A detailed description can be found in later chapters. A functional T cell response requires MHC–peptide binding and a proper interaction of the MHC–peptide ligand with a specific T cell receptor. We need well-characterized data to model the process of binding of peptides to TAP and MHCs which function as T cell epitopes. Some recent investigations include finding and mapping of potential epitopes. Epitope mapping leads to designing effective vaccines. Syfpeithi database [38] (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. IEDB [39] has more than 88382 peptidic epitopes and can be found at http://www.immuneepitope.org/ and ontology-related information (http://ontology.iedb.org/) which has been specifically designed to capture intrinsic, chemical, and biochemical information on immune epitopes and their interactions with molecules of the host immune system. A beta version of IEDB (Immune Epitope Database and Analysis Resource Database) (http://www.immuneepitope.org/) [30], sponsored by the National Institute for Allergy and Infectious Diseases (http://www.niaid.nih.gov) (NIAID), has different tools to find B and T cell epitopes. It had 88382 peptidic epitopes till February 2012. FRED [40] deals with the methods for data processing and to compare the performance of the prediction methods considering experimental values. IMGT® [41] (the international ImMuno GeneTics information system®) (http://www.imgt.org) has a good collection of IG, TR, MHC, and related proteins of the immune system of human and other vertebrates. It has five databases and 15 interactive online tools for sequence, genome, and 3D structure analysis. The IMGT/HLA Database [42] (http://www.ebi.ac.uk/imgt/hla/) provides a specialist database that has 5,518 HLA class I alleles and 1,612 HLA class II alleles. It is a part of the international ImMunoGeneTics project (IMGT).

2.6 Allergy Prediction Databases

Allergy is a steadily increasing health problem for all age groups caused by allergens. Allergens are proteins or glycoproteins recognized by IgE that is produced by the immune system in allergic individuals. 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 [43]. The differences between both kinds of epitopes present in an antigen are as follows: T cell epitopes are only linear (as mentioned earlier) and 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.

Here, we describe allergen prediction databases in brief. One may get details on allergy prediction databases in a later chapter. Allergen Nomenclature database of the International Union of Immunological Societies (IUIS) has allergen database [44] (http://www.allergen.org), which has been last updated in October 2009. AllergenPro database [45] contains 2,434 allergen-related information, e.g., allergens in rice microbes (712 records), animals (617 records), and plants (1,105 records). The web server Allergome 4.0 [46] (www.allergome.org) provides an exhaustive repository of IgE-binding compound data. It has a total of 1,736 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 IT has been applied to allergy data mining. SDAP [47] (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 IgE epitope of allergenic proteins. Its algorithm is based on conserved properties of amino acid side chains. In its latest update, it has 1,478 allergens and isoallergens.

3 Immunomic Tools and Algorithms

The property of an antigen to bind specifically complementary antibodies is known as the antigen's antigenicity; likewise, the ability of an antigen to induce an immune response is called its immunogenicity. The main objective of epitope prediction is to design a molecule that can replace an antigen in the process of either antibody production or antibody detection. Such a molecule can be synthesized or, in case of a protein, its gene can be cloned into an expression vector. Designed molecules are inexpensive and noninfectious in contrast to viruses or bacteria. Epitopes are important for understanding the disease mechanism, host-–pathogen interaction analyses, antimicrobial target discovery, and vaccine design. Traditionally, determination of binding affinity of MHC molecules and antigenic peptides predicts epitopes. The experimental techniques are found to be difficult and time consuming. Due to this reason, several in silico methodologies are being developed and used to identify epitopes. Here, we throw some light on available immunology-related tools and algorithms. These techniques include matrix-driven methods, finding structural binding motifs, quantitative structure–activity relationship (QSAR) analysis, homology modeling, protein threading, docking techniques, and design of several machine-learning algorithms and tools. Table 2 lists some of the tools that deal with B and T cell epitope prediction, allergy prediction, and in silico vaccination. However, detailed description and discussion over the usages of them will be provided in next chapters.

Table 2
Web servers and tools for prediction of B and T cell epitopes, allergens, and in silico vaccination

Web servers and tools	Names	URLs
B cell epitope prediction	ABCpred	http://www.imtech.res.in/raghava/abcpred
	COBEpro	http://www.scartch.proteomics.uci.edu
	Bepipred	http://www.cbs.dtu.dk/services/BepiPred
	IMGT®	http://www.imgt.org
	Bcepred	http://www.imtech.res.in/raghava/bcepred/
	DiscoTope	http://www.cbs.dtu.dk/services/DiscoTope/
	CEP	http://www.115.111.37.205/cgi-bin/cep.pl
	AgAbDb	http://www.115.111.37.206:8080/agabdb2/home.jsp
	MIMOP	Request from franck.molina@cpbs.univ-montp1.fr
	MIMOX	http://www.immunet.cn/mimox/
	Pepitope	http://www.pepitope.tau.ac.il/
	3DEX	http://www.schreiber-abc.com/3dex/
	IEDB	http://www.immuneepitope.org
T cell epitope prediction	MMBPred	http://www.imtech.res.in/raghava/mmbpred/
	NetCTL	http://www.cbs.dtu.dk/services/NetCTL/
	NetMHC 3.0	http://www.cbs.dtu.dk/services/NetMHC/
	TAPPred	http://www.imtech.res.in/raghava/tappred/
	Pcleavage	http://www.imtech.res.in/raghava/pcleavage/
	ElliPro	http://www.tools.immuneepitope.org/tools/ElliPro
	MHCPred	http://www.ddg-pharmfac.net/mhcpred/MHCPred/
	Propred	http://www.imtech.res.in/raghava/propred1/
	EpiToolKit	http://www.epitoolkit.org
	Syfpeithi	http://www.syfpeithi.de
	IMGT®	http://www.imgt.org
	IEDB	http://www.immuneepitope.org/
	EpiJen v 1.0	http://www.ddg-harmfac.net/epijen/EpiJen/EpiJen.htm
Allergy prediction	AlgPred	http://www.imtech.res.in/raghava/algpred
	Allermatch	http://www.allermatch.org
	APPEL	http://www.jing.cz3.nus.edu.sg/cgi-bin/APPEL
	EVALLER	http://www.slv.se/en-gb/Group1/Food-Safety/e-Testing-of-protein-allergenicity/
In silico vaccination	VaxiJen	http://www.ddg-pharmfac.net/vaxijen/
	DyNAVacs	http://www.miracle.igib.res.in/dynavac/
	NERVE	http://www.bio.unipd.it/molbinfo
	VIOLIN	http://www.violinet.org
	Vaxign	http://www.violinet.org/vaxign/

3.1 B Cell Epitope Prediction

Experimental determination of B cell epitopes is time consuming and expensive; there is a need for computational methods for reliable identification of putative B cell epitopes from antigenic sequences. B cell epitopes are antigenic determinants on the surface of pathogens that interact with B cell receptors (BCRs). BCR-binding site is hydrophobic, having six hypervariable loops of

variable length and amino acid composition. B cell epitopes are classified as continuous/linear/sequential and discontinuous/conformational [48]. Linear epitopes are short peptides that correspond to a contiguous amino acid sequence fragment of a protein. However, most epitopes are discontinuous, where distant residues are brought into spatial proximity by protein folding within the folded 3D protein structure. 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 [35, 49].

3.1.1 Prediction of Methodology for Continuous B Cell Epitopes

Methodologies for prediction of continuous B cell epitopes involve sequence-based methods, amino acid propensity scale-based methods, and machine-learning methods.

Sequence-Based Methods

Sequence-based methods generally look for the epitope surface that must be accessible for antibody binding. These methods are limited to the prediction of continuous epitopes. Sequence-based methods have been tested on prediction of two protective epitopes known in influenza A virus hemagglutinin HA1 [50]. The first continuous epitope is the 91–108 epitope (SKAFSNCYPYDVPDYASL), which is a protective epitope in rabbit able to elicit antibodies neutralizing infectivity of influenza viruses [51]. The second continuous epitope is the 127–133 epitope (WTGVTQN) protective against the influenza strain A/Achi/2/68 (H3N2) in mouse [52].

Amino Acid Propensity Scale-Based Methods

Parameters such as hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity, and antigenic propensity of polypeptide chains have been correlated with the location of continuous epitopes. Thus the classical methods of identifying potential linear B cell epitopes from antigenic sequences typically rely on the use of amino acid propensity scales. Amino acid scale-based methods apply amino acid scales to compute the scores of a residue i in a given protein sequence. The i–$(n-1)/2$ neighboring residues on each side of residue i are used to compute the score for residue i in a window of size n. The final score for residue i is the average of the scale values for n amino acids in the window. Pellequer [53] compared several propensity scale methods using a dataset of 14 epitope-annotated proteins. He found that the scales of Parker et al. [54], Chou and Fasman, [55], Levitt [56], and Emini [57] provide better results than the other scales tested [48]. El-Manzalawy et al. [58] compared propensity scale-based methods with a naive Bayes classifier and used two datasets: one is propensity dataset, and the other is from BciPep [35].

Bepitope tool [59] predicts continuous epitopes based on the prediction of protein turns. It is a newer version of PREDITOP [60] and uses more than 30 propensity scale values. Bcepred server [61]

(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 interaction are done on antibody-binding sites on proteins, which help in predicting the linear and conformational B cell epitopes. Taking this into consideration, a database, viz., AgAbDb [62] (http://202.41.70.51:8080/agabdb2/), has been developed which is based on molecular interactions of antigen–antibody cocrystal structures.

Machine-Learning Methods

Machine-learning algorithms and tools are being used to retrieve characteristics of an epitope. Here we describe some of these approaches in brief. Saha and Raghava [63] used feed-forward and recurrent neural networks to predict continuous B cell epitopes in ABCpred (http://www.imtech.res.in/raghava/abcpred). COBEpro [64] is a two-step system for prediction of continuous B cell epitopes. In the first step, COBEpro assigns a fragment epitopic propensity score to protein sequence fragment using SVM. In the second step, it calculates an epitopic propensity score for each residue based on the SVM scores of the peptide fragment in the antigenic sequence. For Bepipred [65], (http://www.cbs.dtu.dk/services/BepiPred), three datasets of linear B cell epitopes were constructed, viz., annotated proteins from literature, AntiJen database [66] (http://www.ddg-pharmfac.net/antijen/AntiJen/antijenhomepage.htm), and Los Alamos HIV database (http://www.hiv.lanl.gov). They tested a number of propensity scale methods on Pellequer dataset [53] and found the best scale by Levitt [56]. Then, they used HMM to predict the location of linear B cell epitopes and tested HMMs on Pellequer dataset to find optimal parameters. HMM was combined with one set of the two best propensity scale methods, i.e., Parker [54] and Levitt [56], to get the more accurate predictions. Currently, ~60–66 % of accuracy has been found for continuous epitope prediction, applying combinations of either amino acid scales or machine-learning techniques. The higher accuracy could possibly be achieved by improving the quality of existing B cell epitope datasets [48].

3.1.2 Prediction Methodology for Discontinuous B Cell Epitopes

The characterization and prediction of B cell epitopes are mainly conformational dependent based on the knowledge of the protein three-dimensional structure; thus the task of prediction is more difficult compared to that of T cell epitopes. Changes in protein folding may lead to changes in the number of epitopes [43]. The most accurate way to identify B cell epitope is through X-ray crystallography. Here we describe some of the prediction methods for conformational B cell epitopes in brief. Anderson et al. presented a method called DiscoTope [67] (http://www.cbs.dtu.dk/services/DiscoTope/), which is a combination of amino acid statistics, spatial information, and surface exposure. It detects 15.5 % of residues

located in discontinuous epitopes with a specificity of 95 %. It is said to be the first method developed for prediction of discontinuous B cell epitope with better performance than methods based only on sequence data. PEPITO [68] uses a weighted linear combination of amino acid propensity scores and half-sphere exposure values [69] which encode side chain orientation and solvent accessibility of amino acid residues for the prediction of conformational epitopes. Authors have also reported its improvement in performance over DiscoTope method.

Bublil et al. developed Mapitope [70] for conformational B cell epitope mapping. The hypothesis behind Mapitope is that the simplest meaningful fragment of an epitope is an amino acid pair (AAP) of residues that lie within the epitope, which are the results 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 1–3 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. [71] 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 (PTM) prediction. By their observation, they found high antigenicity, low variability, and low likelihood of PTM for the identification of biorelevant sites. Ponomarenko [48] assembled non-redundant datasets of repetitive 3D structure of antigen and antigen–antibody complexes from the PDB. CEP web interface [72] (http://www.115.111.37.205/cgi-bin/cep.pl) predicts conformational and sequential epitopes and also antigenic determinants. Less availability of the 3D structure data of protein antigens limits the utility of this server. A recent approach has focused on the impact of interior residues, different contributions of adjacent residues, and imbalanced data which contain much more non-epitope residues than epitope and applied random forest (RF) algorithm for the prediction of conformational B cell epitope prediction [73]. This tool is available at http://www.code.google.com/p/my-project-bpredictor/downloads/list.

Mimotope-Based Methodology

Phage display library is widely used for finding protein–protein interactions (specially in antibody–antigen interactions), protein function identification, and development of new drugs and vaccines [74]. Pizzi et al. [75] have proposed an approach for mapping B cell epitopes, in which a phage display library of random peptides is scanned against a desired antibody to obtain mimotopes that bind to the antibody with high affinity. It is assumed that this panel of mimotopes mimics the physicochemical properties and

spatial organization of the genuine epitopes [34, 74 and 76]. Mimotopes and antigens are both recognized by the same antibody paratope. Mimotopes are said to be the imitated part of the epitope. It is possible that mimotope may have some valuable information about epitope. However, homology may not exist between the mimotope and the epitope of the native antigen. This mimicry exists due to similarities in physiochemical properties and spatial organization [76]. Considering these properties, mimotope pools are being used to mine information to predict an epitope.

Using the above concept, MIMOP tool [76] has been developed. MIMOP predicts linear and conformational epitopes based on two algorithms, viz., MimAlign uses degenerated alignment analyses, and MimCons is based on consensus identification. MIMOX [77] (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, viz., Pepsurf [78] and Mapitope [70]. It maps each mimotope to map them onto the solved structure of antigen surface. Alignment of mimotope is done first in MIMOX, so 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 3D structure. The 3D structure is considered only following the alignment stage. Given the 3D structure of an antigen and a set of mimotopes (or a motif sequence derived from the set of mimotopes), Pep-3D-Search [79] (http://kyc.nenu.edu.cn/Pep3DSearch/) can be used in two modes: mimotope or motif. It can be used for localizing the surface region mimicked by the mimotopes.

Sometimes linear peptides mimic conformational epitopes. The same phage display peptide libraries by screening with the respective antibodies are used to select these mimotopes. Schreiber et al. [80] presented software, 3DEX (3D-Epitope-Explorer) (http://www.schreiber-abc.com/3dex/), that allows localizing linear peptide sequences within 3D structures of proteins. Its algorithm takes into account the physiochemical neighborhood 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 HIV-positive patient plasma within 3D structure of gp120.

Hybrid (Ensemble) Prediction Method

Ensemble methods combine the predictions of several predictors and often outperform individual predictors in many biomolecular sequence and structure classification studies [81]. Several strategies for combining a set of predictors, *S*, into a single consensus or meta-predictor exist: (1) majority voting, (2) weighted linear

combination, and (3) meta-learning [82]. A large number of nearest neighbor- and decision tree-based classifiers are trained using different sets of training data features for developing an ensemble of linear B cell epitope classifiers [83].

3.2 T Cell Epitope Prediction

The current challenge in immunological prediction software is to predict interacting molecules to a high degree of accuracy. The most popular methods currently available are based on binding affinity predictions for a range of MHC molecules. It is necessary to bind antigenic peptides with MHC so that cytotoxic T cells can recognize them. Thus, identification of MHC-binding peptides is a central part of any algorithm which predicts T cell epitopes. There exist several methodologies for prediction of MHC-binding peptides, which are based on the idea of quantitative matrices, hidden Markov model (HMM), artificial neural networks (ANNs), support vector machine (SVM), and structure of the peptides. Here we describe the abovementioned approaches in brief. One may find the details of these methodologies, among others, in later chapters.

3.2.1 Matrix-Driven Methods

Huang and Dai [84] first investigated a new encoding scheme of peptides based on BLOSUM matrix with the amino acid indicator vectors for direct prediction of T cell epitopes. It replaced each nonzero entry in the amino acid indicator vector by the corresponding value appeared in the diagonal entries in BLOSUM matrix. MMBPred [85] (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.

3.2.2 Hidden Markov Model-Based Method

Transfer-associated protein (TAP) 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 (ER). Zhang et al. developed $PRED^{TAP}$ [86] 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. Secondly, they used second-order HMM. The results are both sensitive and specific.

3.2.3 Artificial Neural Network-Based Method

ANNs can identify each amino acid residue and interactions between adjacent ones in a potential epitope. An ANN for a particular MHC molecule is trained to recognize associated input sequence and outputs, viz., the binding affinity for that sequence with the MHC molecule [87]. Trained ANN can predict the binding affinity of novel peptide sequences. Neilson et al. [88] described an improved neural network model to predict T cell class I epitopes.

They combined a sparse encoding, BLOSUM encoding, and input derived from HMM. The dataset consists of 528 nine-mer 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. [89]. NetCTL server [90] (http://www.cbs.dtu.dk/services/NetCTL/) has method to integrate the prediction of peptide MHC class I binding, proteasomal C terminal cleavage, and TAP transport efficiency. NetMHC server 3.0 [91] (http://www.cbs.dtu.dk/services/NetMHC/) uses ANN and weight matrices. It has been trained on data from 55 MHC peptides (43 human and 12 nonhuman) and position-specific scoring matrices (PSSMs) for additional 67 HLA alleles.

Prediction of MHC class II binding peptides is found to be difficult due to the 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. developed PERUN [92], 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 ANNs. They used PlaNet package version 5.6 [93] to design and train a three-layered fully connected feed-forward ANN. The whole process of MHC class I ligands' degradation and presentation has been modeled in EpiJen [94] (http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm) in an integrative approach. It is a multi-step algorithm for T cell epitope prediction, based on quantitative matrices, which belongs to the next generation of in silico T cell epitope identification methods.

3.2.4 Other Machine-Learning Methods

Ant colony search systems (ACSs) have been found useful for solving combinatorial optimization problems and can be applied to the identification of a multiple alignment of a set of peptides. Basically, ACSs [95] attempt to find an optimal alignment for a given set of peptides based on the search strategy. For TAPPred [96] (http://www.imtech.res.in/raghava/tappred/), nine features of amino acids have been analyzed to find the correlation between binding affinity and physiochemical properties. An SVM-based method to predict TAP binding affinity of peptides has been developed 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. Nanni [97] demonstrated the use of SVM and support vector (SV) data description to predict T cell epitope. It is experimentally established that the immunoproteasome is involved in the generation of the MHC class I ligand. For this purpose, Pcleavage [98] (http://www.imtech.res.in/raghava/pcleavage/) has been developed to predict both kinds of cleavage sites in antigenic proteins. It uses SVM [99], Parallel Exemplar based Learning (PEBLS) [100], and Waikato Environment for Knowledge Analysis (Weka) [101].

3.2.5 Structure-Based Prediction

Accurate identification of peptides that bind to specific MHC molecules is important for understanding the underlying mechanism of immune recognition, for developing effective peptide-based vaccines, and for immunotherapies for allergy and autoimmunity. Current methods are mostly based on peptide binding affinity to MHC for predicting T cell epitope. 3D QSAR technology CoMSIA has been applied to the problem of peptide–MHC binding [102]. It uses the interaction potential around aligned sets of 3D peptide structures to describe binding. TEPITOPE [103] is used to predict promiscuous and allele-specific HLA II-restricted T cell epitope in silico. TEPITOPE's user interface has a display and comparison of pocket profiles, and it finds similar HLA II differing in their binding capacity for a given peptide sequence. It can be applied to only 51 out of over 700 known HLA-DR molecules. A new method called as TEPITOPEpan (http://www.biokdd.fudan.edu.cn/Service/TEPITOPEpan/) is developed by extrapolating from the binding specificities of HLA-DR molecules characterized by TEPITOPE to those uncharacterized [104].

T epitope designer [105], a web server, 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. [106] 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 [107] (http://www.tools.immuneepitope.org/tools/ElliPro) is a web tool that implements a modified version of Thornton 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 approximation of the protein shape as an ellipsoid, calculation of the residue protrusion index (PI), and clustering of neighboring residue based on their PI 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. in Edward Jenner Institute for Vaccine Research, UK, introduced MHCPred version 2.0 [108] (http://www.ddg-pharmfac.net/mhcpred/MHCPred/). It is a perl implementation of 2D QSAR application to peptide–MHC prediction and covers both class I and class II MHC allele peptide specificity models. Peptide that can bind to MHC on the tumor cell surface has the potential to initiate a host immune response against the tumor. Schiewe et al. [109] developed an algorithm PeSSI (peptide–MHC prediction of structure through solvated interfaces) for flexible structure prediction of peptide binding to MHC molecule. They used cancer testis (CT) antigens, KU-CT-1, that are potential to bind HLA-A2.

Jojic et al. [110] developed an improved structure-based model which used known 3D structures of a small number of MHC–peptide complexes, MHC class I sequence, known binding energies for MHC–peptide complexes, and larger binary dataset having 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. [111] 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 [112].

Immunodominant peptides are being used for rational design of peptide vaccines focusing on T cell immunity. Altuvia et al. [113] focused on antigenic peptides recognized by cytotoxic T cells. They applied the threading approach to screen a library of peptide sequence and identified the ones that optimally fit within the MHC groove. Propred [114] (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. [115]. EpiToolKit [116] (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.

3.2.6 Molecular Dynamics-Based Prediction

Molecular dynamics (MD) describes single and collective motion of atoms within a molecular system and provides a means by which one can measure theoretically that cannot be measured experimentally [117]. It is particularly suitable for the simulation and analysis of the otherwise inaccessible details of MHC–peptide interaction and of the immune synapse. Zhang et al. [118] were among the first who uses MD as a tool to explore peptide–MHC binding. They focused on docking using MD as well as on calculating free energies. Free energy calculations of the wild-type and the variant human T cell lymphotropic virus type 1 Tax peptide (LLFGYPVYV—wild Tax and LLFGYAVYV—mutant Tax) presented by the MHC to the TCR have been performed using large-scale massively parallel molecular dynamics simulations [119].

3.3 Allergy Informatics

Allergy is caused by adverse immunological reaction, and the causative agents are known as allergens that are otherwise not harmful in nature. An allergen cross-links immunoglobulin E (IgE) antibodies on mast cells or basophils and releases inflammatory mediators that cause allergy symptoms. Biotechnology- and genetic engineering-derived food contains some foreign proteins, which

can be allergic to many human beings. 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 analyze the structural and physiochemical properties of food allergen proteins. Current efforts in allergy informatics are primarily focused on prediction of T and B cell epitopes and assessment of allergenicity.

Allergy occurs by both extrinsic and intrinsic factors. Type I hypersensitive reaction is induced by certain allergens that elicit IgE antibodies [1]. 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 that has at least six same contiguous amino acids or a window of 80 amino acids when compared with known allergens is considered as allergen. It has already been established that allergens do not share common structural characteristics. Thus allergen databases are being used as reference for finding the sequence similarity in allergenicity evaluation [120]. It is said that a protein is considered as an allergen if it has a region or peptides identical to a known IgE epitope.

Allergen prediction method proposed by Kong et al. [121] 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 non-allergen test set from the given reference [122]. They developed a database which has all possible combinations of two motifs from the set of allergenic motifs by using motif length of 35 amino acids and motif number of 500. Zorzet et al. [123] introduced a computational approach for classifying the amino acid sequences in allergens and non-allergens. They identified preprocessed 91 food allergens from various specialized public repositories of food allergy and SWALL database (SWISSPROT and TrEMBL).

AlgPred [124] (http://www.imtech.res.in/raghava/algpred) uses SVM and a similarity-based approach for analysis and scanned all 183 IgE epitopes against all proteins of the dataset. The server allows using a hybrid option to predict allergen using combined approach (SVMc, IgE epitope, ARPs BLAST, and MAST).

Stadler et al. [120] used MEME motif discovery tool to identify the most relevant motif present in 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 (http://www.allermatch.org) is a web tool that uses sliding window approach to predict potential allergenicity of proteins [125].

It is done according to the current recommendations of the FAO/WHO Expert Consultation, [126] as outlined in Codex alimentarius [127]; however, 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.

APPEL [128] (Allergen Protein Prediction E-Lab) (http://www.jing.cz3.nus.edu.sg/cgi-bin/APPEL) 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 statistical method, and it has the potential to discover novel allergen proteins. EVALLER [129] web server (http://www.slv.se/en-gb/Group1/Food-Safety/e-Testing-of-protein-allergenicity/) uses filtered length-adjusted allergen peptides (DFLAP) method [130] (via ulfh@slv.se) to identify the potential allergen proteins. DFLAP extracts variable length allergen sequence fragments and employs SVM.

EVALLER and APPEL servers assigned all calmodulins or calmodulin-like proteins as presumably non-allergens [128, 129]. But a conventional alignment approach (e.g., 35 % similarity over 80 amino acid segments) gives preference to find sequence similarity between input proteins and known allergens and puts abovementioned proteins in allergen category. These proteins are presumable 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 sequence similarity method alone. Details related with allergen prediction approaches may be found in later chapters.

4 Applications of Immunoinformatics

The use of immunological databases and prediction software has become an important part of the scientific research as they allow us to predict the interaction of molecules involved in an immune response, thereby significantly shortening experimental procedure. In this section, we focus on applications of immunoinformatics. It includes in silico vaccine design and immune system modeling and immunoinformatics for cancer diagnosis and therapy. It also explores the idea of integrating systems biology with immunoinformatics.

4.1 Reverse Engineering for Vaccine Design

Vaccines can be live attenuated whole pathogens, subunits, or epitope based. It is possible to design attenuated pathogens by removing virulence factors or reducing their metabolic capacity. These procedures can be done through computational design and discovery. Several in silico techniques have been developed to identify suitable vaccine candidates, principally proteins within

pathogen genomes that have antigenic properties. Generally used vaccines are live attenuated or killed bacteria or viruses (examples include cholera, polio, measles). Thus there is a concern about the safety of these vaccines; if they are incompletely attenuated or killed, they may revert their pathogenicity or cause undesirable immune reactions. On the other hand, synthetic peptides are considered as candidates for safe vaccines. Methods predicting immunogenic peptides could lead to rational vaccine design. Genome sequencing, comparative proteomics, and immunoinformatics tools are well developed to design new vaccines. "Reverse vaccinology," a new concept, analyzes the entire genome to identify potentially antigenic extracellular proteins and thus helps in saving time and money. It was pioneered for *Neisseria meningitides* responsible for sepsis and meningococcal meningitides, and the vaccine type is conjugate based on capsular polysaccharide. These vaccines are available for pathogenic *N. meningitides A, C, Y, and W135* [131].

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 [132]. 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 of the whole-protein vaccines. It may induce immune response against immunodominant epitopes [133]. This kind of vaccine has a single start codon with an epitope which can be inserted consecutively in the construct [134]. The prediction of promiscuous binding ligands is considered to be a prerequisite for the most subunit vaccine design strategies [135]. It is originally named as "reverse immunogenetics" where T cell epitope mapping tools were employed to find new protein candidates for vaccines and diagnostic tests [136]. Epitope-driven vaccine design allows the discovery of previously unknown and undescribed antigens and epitopes as vaccine candidates. The major disadvantage of the epitope-based approach is that algorithms may fail to predict all the relevant epitopes [137]. A web server, PEPVAC (Promiscuous EPitope-based VACcine) (http://immunax.dfci.harvard.edu/PEPVAC/) [138], is optimized for the formulation of multi-epitope vaccines with broad population coverage. This optimization is accomplished through the prediction of peptides that bind to several HLA molecules with similar peptide-binding specificity.

Peptide-based vaccine design: Small peptides derived from epitopes are used as peptide-based vaccines. These peptides are recognized by MHC class I and thus boost the immune response. Three novel classes of methods have been described to predict MHC-binding peptides and a voting scheme to integrate them for improved

results [139]. The first method is based on quadratic programming applied to quantitative and qualitative data. 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: Some proteins have similar structure and biological properties, but they may lack sequence similarity. For these kinds of proteins, a new alignment-free approach for antigen prediction has been proposed, which uses three datasets—each for bacteria, viruses, and tumors [140]. The models were validated using leave-one-out cross-validation (LOO-CV) on the whole sets and by external validation using test sets and were implemented in a server called VaxiJen version 2.0 (http://www.ddg-pharmfac.net/vaxijen/).

DNA vaccines: DNA vaccines produce cell-mediated and humoral immune response and are very useful in defending intracellular pathogens. It uses plasmid DNA, which contains a DNA sequence coding for an antigen and a promoter for gene expression in the mammalian cell. Plasmid DNA does not need a viral vector for delivery. Naked DNA is safe and can be used to sustain the expression of antigen in cells for longer periods of time than RNA or protein vaccines. The DNA delivers antigen as well as activates innate immunity and an adaptive immunity against cancer antigens. DyNAVacs [141] (http://www.miracle.igib.res.in/dynavac/) incorporates different modules like codon optimization for heterologous expression of genes in bacteria, yeast, and plant, mapping restriction enzyme sites, primer design, Kozak sequence insertion, custom sequence insertion, and design of genes for gene therapy.

The crucial question in deciding vaccine protocol is the vaccination schedule, i.e., is to decide whether the chronic protocol is able to give 100 % protection or shorter protocols could be applied. Thus a mathematical model/simulator (SimTriplex) which describes the immune response activated by the triplex vaccine has been developed [142]. Immunological prevention of cancer has been obtained in HER-2/neu transgenic mice using a vaccine that combines three different immune stimuli (triplex vaccine) that is repeatedly administered for the entire life-span of the host (chronic protocol).

The software NERVE [143] (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 task by implementing and integrating six different kinds of analyses. Xiang et al. developed a web-based database system, VIOLIN [144] (Vaccine Investigation and Online Information Network) (http://www.violinet.org), which curates, stores, and analyzes published vaccine data. It contains four integrated literature mining and search programs, viz.,

Litsearch, Vaxpresso, Vaxmesh, and Vaxlert. They have developed a web-based vaccine design system called Vaxign [145], 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. However, synthetic vaccine candidates must be tested experimentally to demonstrate their ability to generate neutralizing antibodies.

4.2 Immune System Modeling

The immune system can be seen as a parallel, information processing system that learns through examples, constantly adapts itself to new situations, and possesses a distributive memory for patterns. For theoretical immunology, immune system models and simulations can describe more insights into various interactions resulting in immunological phenomena. 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. Using visual modeling application described by Gong and Cai [146] one can understand the adaptive immune system effectively. The hierarchical immune system consists of inherent immune tier, adaptive immune tier, and immune cell tier. It is designed and visualized with 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.eu) and Virolab (http://www.virolab.org/) projects are working to simulate immune systems. ImmunoGrid tries to simulate immune processes by combining experiments and computational studies, while Virolab attempts to develop a virtual lab for infectious diseases by examining the genetic causes of human illnesses [132]. SIMISYS 0.3 [147] 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 modeling generate enormous amount of data, but there should be a balance between virtual and real experimental data. Computationally generated data needs to be formally tested and translated into real knowledge. Post-genomic era needs to exchange data from wet lab to simulation and vice versa [148]. 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.

4.3 Immunoinformatics for Cancer Diagnosis and Therapy

Antigen presentation plays a central role in the immune response and as a result also in immunotherapeutic methods like antitumor vaccination. There is a need to rapidly screen the antigens and to design specific types of expression constructs for immunotherapy of cancer. Competent immune responses to cancer are likely to be restricted to the immunome of a specific cancer, including the set of antigens that drive successful immune responses. However, it is still difficult to find the set of antigens that varies between different tumors. Antitumor vaccination takes advantage of in vivo processes, and it harnesses the full power of the immune system, unlike the more artificial ex vivo expansion of T cells [149].

Changes in the cancer diagnosis and prevention are being supported by informatics [150]. For example, the Cancer Biomedical Informatics Grid (caBIG) connects a network of 500 individuals and 50 institutions who share data and analyze tools to speed up the development of innovative approaches for the prevention and treatment of cancer [151]. The 2005 database issue of Nucleic Acids Research lists 14 cancer-related molecular databases, which mainly focus on cancer-related genes and gene expression [152]. Listings of tumor antigens are also available [153]. This list includes antigens that have defined T cell epitopes. Tumor-associated antigens (TAA) have played a vital role in both diagnosis and treatment of human carcinomas, such as prostate-specific antigen (PSA) in the diagnosis of prostate cancer. Despite this, the process of TAA identification has often been hampered by the complicated lab procedures. To fasten the process of tumor antigen discovery, and improve diagnosis and treatment of human carcinoma, a publicly available database Human Potential Tumor Associated Antigen (HPtaa) database (http://www.hptaa.org) has been established [154]. Systems biology approaches target identification of a small number of antigens expressed by cancer cells that are suitable targets of immune responses against cancer. A proteomic mapping of in vivo targets for antibodies in lungs, and solid tumors in experimental animals define aminopeptidase-P and annexin A1 as targets of anticancer immune responses [155]. Informatic methods have also been used for classification of tumors into subtypes, which supports decision making for the selection of therapeutic approaches; however, such applications in cancer immunology are yet to come [156].

Vaccine against tumors: Reliable predictions of immunogenic T cell epitope peptides are crucial for rational vaccine design and represent a key problem in immunoinformatics. Computational approaches have been developed to facilitate the process of epitope detection and show potential applications to the immunotherapeutic treatment of cancer. Epitope-driven vaccine design employs these bioinformatics algorithms to identify potential targets of vaccines against cancer [157]. The development of epitope-based

DNA vaccines and their antitumor effects in preclinical research against B cell lymphoma have been described [158].

Most immunotherapeutic approaches work on the induction of antitumor CD8+ T cells, which exhibit cytolytic activity towards tumor cells expressing tumor-specific or tumor-associated Ags. But the immunization strategies that focus solely on CD8+ T cell immunity might prove to be insufficient because they will be unable to provide long-term protective immunity [159]. It has been shown that the peptides predicted to bind MHC can elicit a tumor-killing cytotoxic T lymphocyte (CTL) response [160]. Although CTLs have been found to be the key player in the generation of antitumor therapeutic effects, sometimes they also remain as suboptimal. CD4+ T cells are critical for the generation and maintenance of CTL response through providing cytokines or by major pathway, i.e., dendritic cell licensing [161, 162]. Class II MHC-bound epitopes activate CD4+ T cells and maintain effective CTL response that plays an important role in the antitumor response [163, 164].

CD4+ T cells determine the functional status of both innate and adaptive immune responses; thus, the inclusion of appropriate CD4+ T cell epitopes may be essential for vaccine efficacy. Idiotypic immunoglobulin M (IgM) expressed by B cell lymphoma is a clonal marker and a tumor-specific antigen. Thus, it can be used as an immune target. Specific immunogenic epitopes identified from these tumor antigens can be used as vaccines to activate an immune response against tumor cells [165]. Concerning to lymphoproliferative malignancies, tetanus toxin fragment C (TTFrC)-fusion vaccine design was able to activate anti-Id antibody responses and to suppress tumor growth in murine models [166, 167] as well as was effective in inducing CD8+ CTL in several tumor models [168].

4.4 Immunoinformatics and Systems Biology for Personalized Medicine

The idea to integrate immunoinformatics with systems biology approaches is for the better understanding of immune-related diseases at various systems levels. This integration can open the path of several translational studies for better clinical practices. The association between a disease and genetic variations is one of the most important aspects in pharmacogenomics and development of personalized medicine. Figure 2 shows the integration that leads to the development of personalized medicine (inspired by 169). The information about allele frequencies of immune molecules in a human population is important as different patient subgroups can be identified with different vaccine or drug responses [169]. For example, an SNP (S427T) in the innate immune gene interferon regulatory factor 3 (IRF3) has been associated with increased risk of human papillomavirus (HPV) persistence and cervical cancer [170]. Genomic variation databases such as HapMap (http://www.snp.cshl.org/) and dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) provide information on individual genotype data. The Allele Frequencies Database can be used to search for polymorphic regions of various populations on histocompatibility and immunogenetics

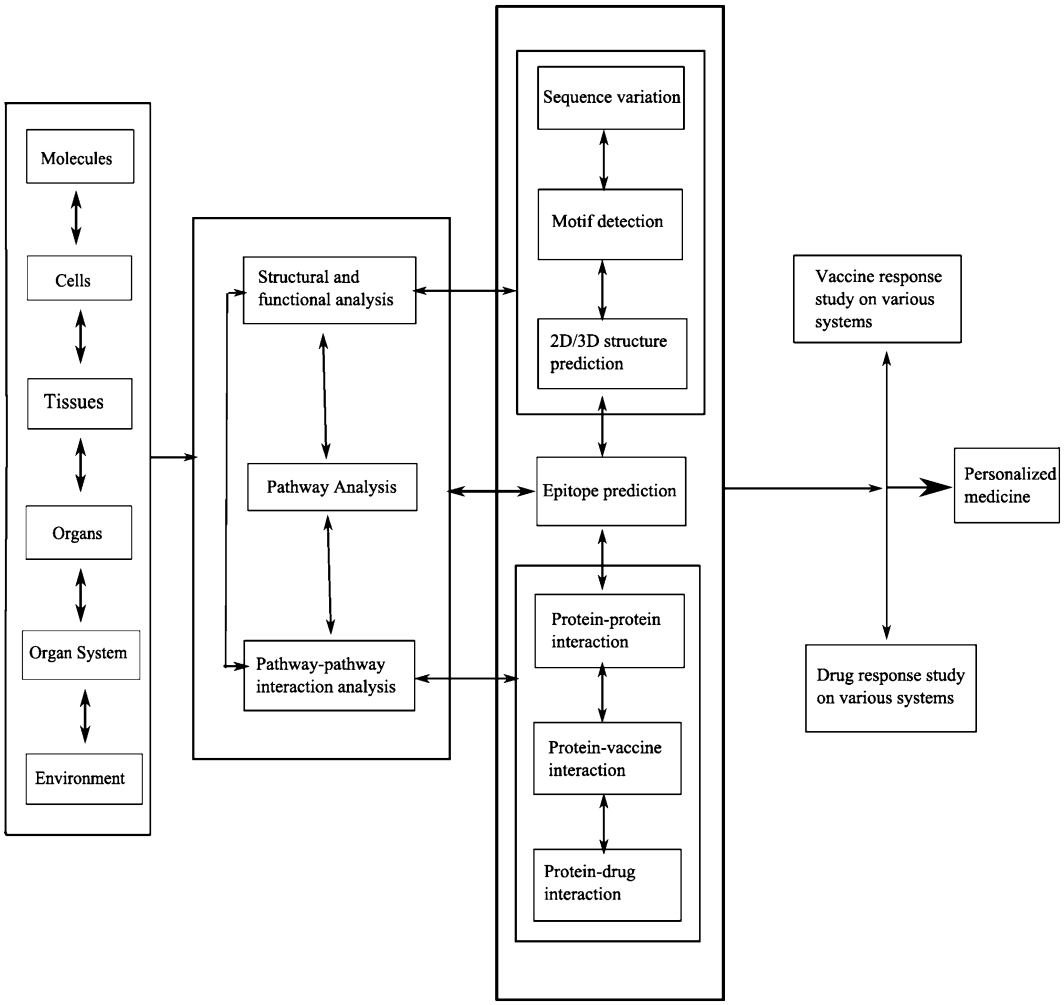

Fig. 2 An integration of immunoinformatics and systems biology, leading to the development of a personalized medicine, inspired by ref. [169]

(http://www.allelefrequencies.net/). This includes polymorphism information on HLA, cytokines, and killer-cell immunoglobulin-like receptors (KIR). Thus, there is a scope for the development of optimized vaccines and drugs tailored to personalized prevention and treatment through the integration of systems biology and immunoinformatics.

5 Conclusions and Discussions

High-throughput experimental techniques are combined with immunoinformatics, resulting in explosive growth of immunology. This is as similar as the event that has transformed genetics into genomics. Immunoinformatics may be placed at the junction point between experimental and computational approaches as it reduces

time and cost involved in traditional study of immunology. This review considers useful online immunological databases, tools, and web servers and explores the application of immunoinformatics in various scientific domains with an emphasis on reverse vaccinology.

Earlier approaches have some limitations in handling real data (nonlinear data). Machine-learning techniques can deal with nonlinear data. SVM (a statistical learning methodology) is a learning technique which supports continuous and categorical variables. SVM is better than ANN, as it attains global minimum and is capable of working with less number of training patterns [171]. Thus both sequence characteristics and computational techniques should be integrated to acquire higher prediction accuracy.

"Reverse vaccinology" is a revolution in immunology as it uses the whole spectrum of antigens. This helps in using pools of vaccine candidates which otherwise would be missed (because of poor or no in vitro experimental information or facing problem in culturing the specific pathogen) [171]. 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 prime target of epitope-based T cell vaccine design.

Immunoinformatics models simulate the real behavior of immune system processes and thus help to get the kinetics of cells during immune responses. It is engineered in such a way that it can be studied and interpreted easily and can be rebuilt if new experimental data are introduced. These mathematical models remove the uncertainty of the systems as they are found to be closed to wet lab experiments. It leads to design the path for refinement and model the new experiments. But they cannot be directly compared to real biological data as they rely on assumptions only. There is no data for extended time spans available to validate the model. This limits the accuracy of the results. Currently models are designed in such a way that they simulate the biological data only over a fixed time period [172]. It should have the ability to show the system's changes over an extended time period for immune response in case of antigen attack or drug administration. This will reduce the necessity of experimental research.

Drug response to a host's immune system can be better studied through computational models. Effect of drug administration can be added to model the immune system to find the drug efficacy [172]. Immune system/drug response study provides an idea about the dose composition, drug dosage duration, age of the patient, and other parameters. These modeling capabilities may lead to designing a drug, which can treat a disease without any side effects. Thus the idea of integrating systems biology with immunoinformatics can lead to better clinical trials.

Acknowledgment

Ms. Namrata Tomar, one of the authors, gratefully acknowledges CSIR, India, for providing her a Senior Research Fellowship (9/93(0145)/12, EMR-I).

References

1. Thomas K, Goldsby J, Osborne RA, Barbara A, Kuby J (2006) Kuby immunology, 6th edn. Freeman and Co., WH
2. Kimbrell DA, Beutler B (2001) The evolution and genetics of innate immunity. Nat Rev Genet 2:256–267
3. Korber B, LaBute M, Yusim K (2006) Immunoinformatics: comes of Age. PLoS Comput Biol 2:0484–0492
4. Gardy JL, Lynn DJ, Brinkman FSL, Rew H (2009) Enabling a systems biology approach to immunology: focus on innate immunity. Trends Immunol 30:249–262
5. Davies MN, Flower DR (2007) Harnessing bioinformatics to discover new vaccine. Drug Discov Today 12:389–395
6. Ortutay C, Vihinen M (2009) Immunome Knowledge base (IKB): An integrated service for immunome research. BMC Immunol 10
7. Sette A, Fleri W, Peters B, Sathiamurthy M, Bui HH (2005) A roadmap for the immunomics of category A-C pathogens. Immunity 22: 155–161
8. De Groot AS (2006) Immunomics: discovering new targets for vaccine and therapeutics. Drug Discov Today 11:203–209
9. Grainger DJ (2004) Immunomics: principles and practice. IRTL 2:1–6
10. No K, Everse J, Je D, Fe S, Cy L, Clt L, Ss T, Mosbach K (1974) Purification and separation of pyridine nucleotide-linked dehydrogenases by affinity chromatography techniques. Proc Natl Acad Sci U S A 71:3450–3454
11. Davey HM (2004) Flow cytometric techniques for the detection of microorganisms. Methods Cell Sci 24:91–97
12. Durkin MM, Connolly PA, Wheat LJ (1997) Comparison of radioimmunoassay and enzyme-linked immunoassay methods for detection of *histoplasma capsulatum* var. *capsulatum* antigen. J Clin Microbiol 35:2252–2255
13. Ma H, Shieh KJ, Lee SL (2006) Study of ELISA technique. Nature 4:36–37
14. Levine MA, Thornton P, Forman SJ, Hale PV, Holdorf D, Rouault CL, Powars D, Feinstein DI, Lukes RJ (1980) Positive Coombs test in Hodgkin's disease: significance and implications. Blood 55:607–611
15. Nishimaki T, Sagawa K, Motogi S, Saito K, Morito T, Yoshida H, Kasukawa R (1987) A competitive inhibition test of enzyme immunoassay for the anti-nRNP antibody. J Immunol Methods 100:157–160
16. Wanga B, Huaa RH, Tiana Z-J, Chena N-S, Zhaoa F-R, Liua T-Q, Wanga Y-F, Tong G-Z (2009) Identification of a virus-specific and conserved B-cell epitope on NS1 protein of Japanese encephalitis virus. Virus Res 141: 90–95
17. Admon A, Barnea E, Ziv T (2003) Tumor antigens and proteomics from the point of view of the major histocompatibility complex peptides. Mol Cell Proteomics 2:388–398
18. Boon T, Coulie PG, Eynde den BV (1997) Tumor antigens recognized by T cells. Immunol Today 18:267–268
19. De Groot AS, Sbai H, Aubin CS, Mcmurry J, Martin W (2002) Immuno-informatics: mining genomes for vaccine components. Immunol Cell Biol 80:225–269
20. Quintana FJ, Hagedorn PH, Gad E, Yifat M, Eutan D, Cohen IR (2004) Functional immunomics: microarray analysis of IgG autoantibody repertoires predicts the future response of mice to induced diabetes. Proc Natl Acad Sci 101:14615–14621
21. Sampson HA (2005) Food allergy-accurately identifying clinical reactivity. Allergy 60:19–24
22. de Vegvar HEN, Robinson WH (2004) Microarray profiling of antiviral antibodies for the development of diagnostics, vaccines, and therapeutics. J Clin Immunol 111: 196–201
23. Neuman de Vegvar HE, Amara RR, Steinman L, Utz PJ, Robinson HL, Robinson WH (2003) Microarray profiling of antibody responses against simian-human immunodeficiency virus: post challenge convergence of reactivities independent of host histocompatibility type and vaccine regimen. J Virol 77: 11125–11138

24. Wang Y (2004) Immunostaining with dissociable antibody microarrays. Proteomics 4:20–26
25. Magdalena J, Odling J, Qiang PH, Martenn S, Joakin L, Uhlen M, Hammarstrom L, Nilsson P (2005) Serum microarrays for large scale screening of protein levels. Mol Cell Proteomics 4:1942–1947
26. Sahin U, Tureci O, Pfreundschuh M (1997) Serological identification of human tumor antigens. Curr Opin Immunol 9:709–716
27. Oelke M, Maus MV, Didiano D, June CH, Mackensen A, Schneck JP (2003) *Ex vivo* induction and expansion of antigen-specific cytotoxic T cells by HLA-Ig coated artificial antigen-presenting cells. Nat Med 9:619–624
28. Braga-Neto UM, Marques ETA (2006) From functional genomics to functional immunomics: new challenges, Old problems, Big rewards. PLoS Comput Biol 2:651–662
29. Nahtman T, Jernberg A, Mahdavifar S, Zerweck J, Schutkowski M, Maeurer M, Reilly M (2007) Validation of peptide epitope microarray experiments and extraction of quality data. J Immunol Methods 328:1–13
30. Peters B, Sidney J, Bourne P et al (2005) The immune epitope database and analysis resource: from vision to blueprint. PLoS Biol 3: 1361–1370
31. Lynn DJ, Winsor GL, Chan C et al (2008) InnateDB: facilitating systems-level analyses of the mammalian innate immune response. Mol Syst Biol 4:1–11
32. Barsky S, Gardy JL, Hancock R, Munzer T (2007) Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. Bioinformatics 23:1040–1042
33. Shanon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504
34. Evans MC (2008) Recent advances in immunoinformatics: application of *in silico* tools to drug development. Curr Opin Drug Discov Devel 11:233–241
35. Saha S, Bhasin M, Raghava GPS (2005) Bcipep: a database of B-cell epitopes. BMC Genomics 6
36. Huang J, Honda W (2006) CED: a conformational epitope. BMC Immunol 7:7
37. Schlessinger A, Ofran Y, Yachdav G, Rost B (2006) Epitome: database of structure-inferred antigenic epitopes. Nucleic Acids Res 34: D777–D780
38. Rammensee HG, Bachmann J, Emmerich NPN, Bachor OA, Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50:213–219
39. Sathiamurthy M, Peters B, Bui HH et al (2005) An ontology for immune epitopes: application to the design of a broad scope database of immune reactivities. Immunome Res 1
40. Feldhahn M, Donnes P, Thiel P, Kohlbacher O (2009) FRED-a framework for T-cell epitope detection. Bioinformatics 25:2758–2759
41. Lefranc M-P, Giudicelli V, Ginestoux C et al (2009) IMGT®, the international ImMuno GeneTics information system®. Nucleic Acids Res 37:D1006–D1012
42. Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SGE (2011) The IMGT/HLA database. Nucleic Acids Res 39(Suppl 1):D1171–D1176
43. Pomes A (2010) Relevant B cell epitopes in allergic disease. Int Arch Allergy Immunol 152:1–11
44. Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TAE, Thomas W (1994) Allergen nomenclature. Bull World Health Organ 72:796–806
45. Kim C, Kwon S, Lee G, Lee H, Choi J, Kim Y, Hahn J (2009) A database for allergenic proteins and tools for allergenicity prediction. Bioinformation 3:344–345
46. Mari A, Scalab E, Palazzob P, Ridolfib S, Zennarob D, Carabella G (2006) Bioinformatics applied to allergy: Allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. Cell Immunol 244:97–100
47. Ivanciuc O, Schein CH, Braun W (2003) SDAP: database and computational tools for allergenic proteins. Nucleic Acids Res 31: 359–362
48. Greenbaum JA, Andersen PH, Blythe M et al (2007) Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. J Mol Recognit 20: 75–82
49. Tong JC, Ren EC (2009) Immunoinformatics: current trends and future directions. Drug Discov Today 14:684–689
50. Bui HH, Peters B, Assarsson E, Mbawuike I, Sette A (2007) Ab and T cell epitopes of influenza A virus, knowledge and opportunities. Proc Natl Acad Sci U S A 104:246–251
51. Muller GM, Shapira M, Arnon R (1982) Anti-influenza response achieved by immunization with a synthetic conjugate. Proc Natl Acad Sci U S A 79:569–573
52. Naruse H, Ogasawara K, Kaneda R, Hatakeyama S, Itoh T, Kida H, Miyazaki T, Good RA, Onoe K (1994) A potential peptide vaccine against two different strains of influenza virus isolated at intervals of about 10 years. Proc Natl Acad Sci U S A 91:9588–9592

53. Pellequer J, Westhof E, Regenmortel MV (1991) Predicting the location of structure of continuous epitopes in proteins from their primary structure. Methods Enzymol 203: 176–201
54. Parker J, Guo D, Hodges R (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
55. Chou PY, Fasman GD (1978) Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol 47:45–148
56. Levitt M (1978) Conformational preferences of amino acids in globular proteins. Biochemistry 17:4277–4285
57. Emini E, Hughes J, Perlow D, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus specific synthetic peptide. J Virol 55:836–839
58. EL-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting protective linear b-cell epitopes using evolutionary information. In: IEEE International Conference on Bioinformatics and Biomedicine 289–292
59. Odorico M, Pellequer JL (2003) BEPITOPE: predicting the location of continuous epitopes and patterns in protein. J Mol Recognit 16: 20–22
60. Pellequer JL, Westhof E (1993) PREDITOP: a program for antigenicity predictions. J Mol Graph 11:204–210
61. Saha S, Raghava GPS (2004) BcePred: Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-chemical Properties. In: Nicosia G, Cutello V, Bentley PJ, Timis J (eds.) ICARIS Springer, LNCS 3239:197–204
62. Ghate AD, Bhagwat BU, Bhosle SG, Gadepalli SM, Kulkarni-Kale UD (2007) Characterization of antibody-binding sites on proteins: development of a knowledgebase and its applications in improving epitope prediction. Protein Pept Lett 14:531–535
63. Saha S, Raghava GPS (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65:40–48
64. Sweredoski MJ, Baldi P (2009) COBEpro: a novel system for predicting continuous B-cell epitopes. Protein Eng Des Sel 22:113–120
65. Larsen JEP, Lund O, Nielsen M (2006) Improved method for predicting linear B cell epitopes. Immunome Res 2
66. Toseland CP, Clayton DJ, McSparron H et al (2005) AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. Immunome Res 1
67. Anderson P, Nielsen M, Lund O (2006) Prediction of residues in discontinuous B cell epitopes using protein 3D structures. Protein Sci 15:2558–2567
68. Sweredoski M, Baldi P (2008) PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. Bioinformatics 24: 1459–1460
69. Hamelryck T (2005) An amino acid has two sides: a new 2D measure provides a different view of solvent exposure. Proteins 59: 38–48
70. Bublil EM, Mayrose NTFI, Penn O, Berman AR (2007) Stepwise prediction of conformational discontinuous B-cell epitopes using the mapitope algorithm. Proteins 68:294–304
71. Sollner J, Grohmann R, Rapberger R, Perco P, Lukas A, Mayer B (2008) Analysis and prediction of protective continuous B cell epitopes on pathogen proteins. Immunome Res 4
72. Kale KU, Bhosle S, Kolaskar AS (2005) CEP: a conformational epitope prediction server. Nucleic Acids Res 33:W168–W171
73. Zhang W, Xiong Y, Zhao M, Zou H, Ye X, Liu J (2011) Prediction of conformational B-cell epitopes from 3D structures by random forests with a distance-based feature. BMC Bioinformatics 12:341
74. Mayrose I, Penn O, Erez E et al (2007) Pepitope: Epitope mapping from affinity-selected peptides. Bioinformatics 23:3244–3246
75. Pizzi E, Cortese R, Tramontano A (1995) Mapping epitopes on protein surfaces. Biopolymers 36:675–680
76. Moreau V, Granier C, Villard S, Laune D, Molina F (2006) Discontinuous epitope prediction based on mimotope analysis. Bioinformatics 22:1088–1095
77. Huang J, Gutteridge A, Honda W, Kanehisa M (2006) MIMOX: a web tool for phage display based epitope mapping. BMC Bioinformatics 7
78. Mayrose I, Shlomi T, Rubinstein ND, Gershoni JM, Ruppin E, Sharan R, Pupko T (2007) Epitope mapping using combinatorial phage-display libraries: a graph-based algorithm. Nucleic Acids Res 35:69–78
79. Huang YX, Bao YL, Guo SY, Wang Y, Zhou CG, Li YX (2008) Pep-3D-Search: a method for B-cell epitope prediction based on mimotope analysis. BMC Bioinformatics 9:538
80. Schreiber A, Humbert M, Benz A, Dietrich U (2005) 3D-Epitope-Explorer (3DEX): Localization of conformational epitopes within

three-dimensional structures of proteins. J Comput Chem 26:879–887

81. Caragea C, Sinapov J, Silvescu A, Dobbs D, Honavar V (2007) Glycosylation site prediction using ensembles of support vector machine classifiers. BMC Bioinformatics 8:438
82. EL-Manzalawy Y, Honavar V (2010) Recent advances in B-cell epitope prediction methods. Immunome Res 6(Suppl 2):2
83. 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
84. Huang L, Dai Y (2006) Direct prediction of T-cell epitopes using support vector machines with novel sequence encoding schemes. J Bioinform Comput Biol 4:93–107
85. Bhasin M, Raghava GPS (2003) Prediction of promiscuous and high-affinity mutated MHC binders. Hybrid Hybridomics 22:229–234
86. Zhang GL, Petrovsky N, Kwoh CK, August JT, Brusic V (2006) Pred^{TAP}: a system for prediction of peptide binding to the human transporter associated with antigen processing. Immunome Res 2
87. Buus S, Lauemoller SL, Worning P, Kesmir C, Frimurer T, Corbet S, Fomsgaard A, Hilden J, Holm A, Brunak S (2003) Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. Tissue Antigens 62: 378–384
88. Neilsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, Brunak S, Lund O (2003) Reliable prediction of T-cell epitopes using networks with novel sequence representations. Protein Sci 12: 1007–1017
89. Buus S, Stryhn A, Winther K, Kirkby N, Pedersen LO (1995) Receptor–ligand interactions measured by an improved spun column chromatography technique. A high efficiency and high throughput size separation method. Biochim Biophys Acta 1243:453–460
90. Larsen MV, Lundegaard C, Lamberth K, Buss S, Lund O, Nielsen M (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics 8
91. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M (2008) 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 36:W509–W512
92. Brusic V, Rudy G, Honeyman M, Hammer J, Harrison L (1998) Prediction of MHC class II-binding peptides using an evolutionary and artificial neural network. Bioinformatics 14: 121–130
93. Miyata J (1991) A User's Guide to PlaNet Version 5.6.
94. Doytchinova IA, Guan P, Flower DR (2006) EpiJen: a server for multistep T cell epitope prediction. BMC Bioinformatics 7:131
95. Dorigo M, Maniezzo V, Colorni A (1996) Ant system: optimization by a colony of cooperating agents. IEEE Trans Syst Man Cybern B 26:29–41
96. Bhasin M, Raghava GPS (2004) Analysis and prediction of affinity of TAP binding peptides using cascade SVM. Protein Sci 13: 596–607
97. Nanni L (2006) Machine learning algorithms for T-cell epitopes prediction. Neurocomputing 69:866–868
98. Bhasin M, Raghava GPS (2005) Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. Nucleic Acids Res 33:W202–W207
99. Joachims T (1999) Marking large-scale support vector machine learning practical. In: Advances in Kernel methods: support vector learning. MIT Press, Cambridge, MA, pp 169–184
100. Cost S, Salzberg S (1993) A weighted nearest neighbor algorithm for learning with symbolic features. Mach Learn 10:57–78
101. Witten IH, Frank E (1999) Data mining: practical machine learning tools and techniques with java implementations, 2nd edn. Morgan Kaufman, San Francisco
102. Flower DR (2003) Towards *in silico* prediction of immunogenic epitopes. Trends Immunol 24:667–674
103. Bian H, Hammer H (2004) Discovery of promiscuous HLA restricted T cell epitope with TEPITOPE. Methods 34:468–475
104. Zhang L, Chen Y, Wong HS, Zhou S, Mamitsuka H, Zhu S (2012) TEPITOPEpan: extending TEPITOPE for peptide binding prediction covering over 700 HLA-DR molecules. PLoS One 7:e30483
105. Kangueane P, Sakharkar MK (2005) T epitope designer: HLA peptide binding prediction server. Bioinformation 1:21–24
106. Zhao B, Mathura VS, Ganapathy R, Moochhala S, Sakharkar MK, Kangneane P (2003) A novel MHCp binding prediction model. Hum Immunol 64:1123–1143
107. Ponomarenko JV, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, Peters B (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics 9

108. Guan P, Doytchinova IA, Zygouri C, Flower DR (2003) MHCPred: a server for quantitative prediction of peptide-MHC binding. Nucleic Acids Res 31:3621–3624
109. Schiewe AJ, Haworth IS (2007) Structure based prediction of MHC-peptide association: algorithm comparison and approach to cancer vaccine design. J Mol Graph Model 26:667–675
110. Jojic N, Gomez MR, Heckerman D, Kadle C, Furman OS (2006) Learning MHC-I peptide binding. Bioinformatics 22:e227–e235
111. Furman OS, Altuvia Y, Sette A, Margalit H (2000) Structure-based prediction of binding peptides to MHC class I molecules: Application to a broad range of MHC alleles. Protein Sci 9:1838–1846
112. Miyazawa S, Jernigan RL (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 256:623–644
113. Altuvia Y, Margalit H (2004) A structure-based approach for prediction of MHC-binding peptides. Methods 34:454–459
114. Singh H, Raghava GPS (2001) Propred: prediction of HLA-DR binding sites. Trends Immunol 17:1236–1237
115. Sturniolo T, Bono E, Ding J et al (1999) Generation of tissue-specific and promiscuous HLA ligand database using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol 17:555–561
116. Feldhahn M, Thiel P, Schuler MM, Hillen N, Stevanovic S, Rammensee HG, Ohlbacher O (2008) EpiToolKit-a web server for computational immunomics. Nucleic Acids Res 1:W519–W522
117. Flower DR, Phadwal K, Macdonald IK, Coveney PV, Davies MN, Wan S (2010) T-cell epitope prediction and immune complex simulation using molecular dynamics: state of the art and persisting challenges. Immunome Res 6(Suppl 2):S4
118. Zhang C, Anderson A (1998) DeLisi C: structural principles that govern the peptide-binding motifs of class I MHC molecules. J Mol Biol 281:929–947
119. Wan S, Coveney PV, Flower DR (2005) Molecular basis of peptide recognition by the TCR: affinity differences calculated using large scale computing. J Immunol 175: 1715–1723
120. Stadler MB, Stadler BM (2003) Allergenicity prediction by protein sequence. FASEB J 17: 1141–1143
121. Kong W, Tan TS, Tham L, Choo KW (2006) Improved prediction of allergenicity by combination of multiple sequence motifs. In Silico Biol 7:77–86
122. Bjorklund AK, Atmadja SD, Zorzet A, Hammerling U, Gustafsson MG (2005) Supervised identification of allergen-representative peptides for *in silico* detection of potentially allergenic proteins. Bioinformatics 21:39–50
123. Zorzet A, Gustafsson M, Hammerling U (2002) Prediction of food protein allergenicity: a bioinformatic learning systems approach. In Silico Biol 2:525–534
124. Saha S, Raghava GPS (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res 34: W202–W209
125. Fiers MWEJ, Kleter GA, Nijland H, Peijnenburg AACM, Peter NJ, Ham RCHJV (2004) Allermatch^{TM}, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. BMC Bioinformatics 5
126. FAO/WHO: Allergenicity of Genetically Modified Foods. http://www.who.int/foodsafety/ publications/biotech/en/ec_jan 2001.pdf, 2001.
127. FAO/WHO: Codex Principles and Guidelines on Foods Derived from Biotechnology ftp://ftp.fao.org/codex/standard/en/CodexTextsBiotechFoods.pdf, 2003.
128. Cui J, Han LY, Li H, Ung CY, Tang ZQ, Zheng CJ, Cao ZW, Chen YZ (2007) Computer prediction of allergen proteins from sequence-derived protein structural and physicochemical properties. Mol Immunol 44:514–520
129. Barrio AM, Atmadja DS, Nistr A, Gustafsson MG, Hammerling U, Rudloff EB (2007) EVALLER: a web server for *in silico* assessment of potential protein allergenicity. Nucleic Acids Res 35:694–700
130. Soeria-Atmadja D, Lundell T, Gustafsson MG, Hammerling U (2006) Computational detection of allergenic proteins attains a new level of accuracy with *in silico* variable-length peptide extraction and machine learning. Nucleic Acids Res 34:3779–3793
131. Pizza M, Scarlato V, Masignani V et al (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287:1816–1820
132. De Groot AS, Rappuoli R (2003) Genome derived vaccines. Expert Rev Vaccines 3: 59–76

133. Gallimore A, Hengartner H, Zinkernagel R (1998) Hierarchies of antigen-specific cytotoxic T cell responses. Immunol Rev 164: 29–36
134. Morris S, Kelly C, Howard A, Li X, Collins F (2000) The immunogenicity of single and combination DNA vaccines against tuberculosis. Vaccine 18:2155–2163
135. Zhao B, Sakharkar KR, Lim CS, Kangueane P, Sakharkar MK (2007) MHC-peptide binding prediction for epitope based vaccine design. Int J Integr Biol 1:127–140
136. Davenport MP, Hill AV (1996) Reverse immunogenetics: from HLA disease associations to vaccine candidates. Mol Med Today 2:38–45
137. Iwai LK, Yoshida M, Sidney J et al (2003) *In silico* prediction of peptides binding to multiple HLA-DR molecules accurately identifies immunodominant epitopes from gp43 of *Paracoccidioides brasiliensis* frequently recognized in primary peripheral blood mononuclear cell responses from sensitized individuals. Mol Med 9:209–219
138. Reche PA, Reinherz EL (2005) PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands. Nucleic Acids Res 33: W138–W142
139. Florea L, Haldorsson B, Kohlbacher O, Schwarty R, Hoffman S, Istrail S (2003) Epitope prediction algorithm for peptide-based vaccine design. Proc IEEE Comput Soc Bioinform Conf 2:17–26
140. Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumor antigens and subunit vaccines. BMC Bioinformatics 8
141. Nagarajan H, Gupta R, Agarwal P, Scaria V, Pillai B (2006) DyNAVacS: an integrative tool for optimized DNA vaccine design. Nucleic Acids Res 34:W264–W266
142. Lollini PL, Motta S, Pappalardo F (2006) Discovery of cancer vaccination protocols with a genetic algorithm driving an agent based simulator. BMC Bioinformatics 7:352
143. Vivona S, Bernante F, Filippini F (2006) NERVE: New enhanced reverse vaccinology environment. BMC Biotechnol 6
144. Xiang Z, Todd T, Ku KP et al (2008) VIOLIN: vaccine investigation and online information network. Nucleic Acids Res 36: 923–928
145. Xianga Z, He Y (2009) Vaxign: a web-based vaccine target design program for reverse vaccinology. Procedia in Vaccinol 1:23–29
146. Gong T, Cai Z (2005) Visual Modeling and Simulation of Adaptive Immune System. In: Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China 6:6116–6119
147. Kalita JK, Chandrashekar K, Hans R, Selvam P, Newell MK (2006) Computational modelling and simulation of the immune system. Int J Bioinform Res Appl 2:63–88
148. Castiglione F, Liso A (2005) The role of computational models of the immune system in designing vaccination strategies. Immunopharmacol Immunotoxicol 27:417–432
149. DeLuca DS, Blasczyk R (2007) The immunoinformatics of cancer immunotherapy. Tissue Antigens 70:265–271
150. Hu H, Brzeski H, Hutchins J et al (2004) Biomedical informatics: development of a comprehensive data warehouse for clinical and genomic breast cancer research. Pharmacogenomics 5:933–941
151. Sanchez W, Gilman B, Kher M, Lagou S, Covitz P (2004) caGRID White Paper (cancer biomedical informatics grid prototype project). National Cancer Institute Center for Bioinformatics (NCICB), USA
152. Galperin MY (2005) The molecular biology database collection: 2005 update. Nucleic Acids Res 33:D5–D24
153. Novellino L, Castelli C, Parmiani G (2005) A listing of human tumor antigens recognized by T-cells: March 2004 update. Cancer Immunol Immunother 54:187–207
154. Wang XS, Zhao HT, Xu QW et al (2006) HPtaa database-potential target genes for clinical diagnosis and immunotherapy of human carcinoma. Nucleic Acids Res 1:D607–D612
155. Oh P, Li Y, Yu J, Durr E, Krasinska KM, Carver LA, Testa JE, Schnitzer JE (2004) Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy. Nature 429:629–635
156. Camp RL, Dolled-Filhart M, Rimm DL (2004) X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res 10: 7252–7259
157. Rosa DS, Ribeiro SP, Cunha-Neto E (2010) CD4+ T cell epitope discovery and rational vaccine design. Arch Immunol Ther Exp 58:121–130
158. Iurescia S, Fioretti D, Fazio VM, Rinaldi M (2012) Epitope-driven DNA vaccine design employing immunoinformatics against B-cell lymphoma: a biotech's challenge. Biotechnol Adv 30:372–383

159. Khanolkar A, Badovinac VP, Harty JT (2007) CD8 T cell memory development: CD4 T cell help is appreciated. Immunol Res 39:94–104
160. Lu J, Celis E (2000) Use of two predictive algorithms of the world wide web for the identification of tumor-reactive T-cell epitopes. Cancer Res 60:5223–5227
161. Smith CM, Wilson NS, Waithman J et al (2004) Cognate CD4(+) T cell licensing of dendritic cells in CD8(+) T cell immunity. Nat Immunol 5:1143–1148
162. Wan YY, Flavell RA (2009) How diverse–CD4 effector T cells and their functions. J Mol Cell Biol 1:20–36
163. Hung K, Hayashi R, Lafond-Walker A, Lowenstein C, Pardoll D, Levitsky H (1998) The central role of CD4+ T cells in the antitumor immune response. J Exp Med 188: 2357–2368
164. Kalams SA, Walker BD (1998) The critical need for CD4 help in maintaining effective cytotoxic T lymphocyte responses. J Exp Med 188:2199–2204
165. Houot R, Levy R (2009) Vaccines for lymphomas: idiotype vaccines and beyond. Blood Rev 23:137–142
166. King CA, Spellerberg MB, Zhu D et al (1998) DNA vaccines with single-chain Fv fused to fragment C of tetanus toxin induce protective immunity against lymphoma and myeloma. Nat Med 4:1281–1286
167. Thirdborough SM, Radcliffe JN, Friedmann PS, Stevenson FK (2002) Vaccination with DNA encoding a single-chain TCR fusion protein induces anticlonotypic immunity and protects against T-cell lymphoma. Cancer Res 62:1757–1760
168. Rice J, Elliott T, Buchan S, Stevenson FK (2001) DNA fusion vaccine designed to induce cytotoxic T cell responses against defined peptide motifs: implications for cancer vaccines. J Immunol 167:1558–1565
169. Yan Q (2010) Immunoinformatics and systems biology methods for personalized medicine. Methods Mol Biol 662:203–220
170. Wang SS, Bratti MC, Rodriguez AC et al (2009) Common variants in immune and DNA repair genes and risk for human papillomavirus persistence and progression to cervical cancer. J Infect Dis 199:20–30
171. Vivona S, Gardy JL, Ramachandran S, Brinkman FSL, Raghava GPS, Flower DR, Filippini F (2008) Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. Trends Biotechnol 26:190–200
172. Daz P, Gillespie M, Krueger J, Prez J, Radebaugh A, Shearman T, Vo G, Wheatley C (2008) A mathematical model of the immune system's response in obesity-related chronic inflammation. McNair/MAOP Summer Research Symposium, Virginia Tech, Blacksburg VA 2:26–4.