We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Databases and Algorithms in Allergen Informatics

Kiran Kadam, Sangeeta Sawant, V.K. Jayaraman and
Urmila Kulkarni-Kale

Additional information is available at the end of the chapter

Abstract

Allergic diseases are considered as one of the major health problems worldwide due to their increasing prevalence. Advancements in genomic, proteomic, and analytical techniques have resulted in considerable progress in the field of allergology, which has led to accumulation of huge amount of data. Allergen bioinformatics comprises allergen-related data resources and computational methods/tools, which deal with an efficient archival, management, and analysis of allergological data. Significant work has been done in the area of allergen bioinformatics that has proven pivotal for the development and progress of this field. In this chapter, we describe the current status of databases and algorithms, encompassing the field of allergen bioinformatics by examining work carried out thus far with respect to features such as allergens and allergenicity, allergen databases, algorithms/tools for allergen/allergenicity prediction, allergen epitope prediction, and allergenic cross-reactivity assessment. This chapter illustrates concepts and algorithms in allergen bioinformatics, as well as it outlines the key areas for potential development in allergology field.

Keywords: allergen databases, allergen epitope prediction, allergenicity, allergenic proteins, bioinformatics

1. Introduction

The immune system represents a very complex system comprising numerous biological molecules and processes, which combine to form body's defense against infectious agents and other threats. Immunity is basically divided into two types such as innate immunity and adaptive immunity [1]. Innate immunity also referred to as natural, native, or nonspecific immunity acts as a first line of defense against common harmful agents. Innate immune response provides

immediate protection and involves number of components such as monocytes, macrophages, neutrophils, cytokines, complement, and epithelial barriers. Adaptive or acquired immunity comprises highly specific immune responses that are elicited against particular pathogens or antigens. These immune responses are either cell mediated or antibody mediated (humoral) and executed by specialized lymphocytes or immunoglobulins, respectively. On certain occasions, the immune system produces immune responses that are harmful for the host organism. Autoimmunity denotes one such case wherein the body elicits immune responses against its own cells and tissues (self-antigens) which lead to development of autoimmune diseases. In some cases, immune system produces inappropriate immune responses known as hypersensitivity, which has deleterious effects on the host organism. Hypersensitivity reactions are categorized into four groups based on the type of immune response and the effector mechanism involved. These are (i) immediate hypersensitivity (type I), (ii) antibody-mediated hypersensitivity (type II), (iii) immune complex–mediated hypersensitivity (type III), and (iv) cell-mediated hypersensitivity (type IV) [2].

Allergic reactions are type I hypersensitivity reactions, which are characterized by induction of specific class of antibodies known as immunoglobulin E (IgE). These reactions are elicited against specific type of antigens commonly referred to as allergens. An allergic reaction involves specialized cells and specific molecules of the immune system [3]. IgE antibodies induced by allergens upon allergic sensitization bind to effector cells such as basophils and mast cells via specific Fc receptors present on the surfaces of those cells. Subsequent exposure to the allergen causes cross-linking of membrane-bound IgE on these effector cells, which leads to their degranulation and release of pharmacologically active agents such as histamine. These pharmacological mediators are responsible for clinical manifestations of allergic reactions in the affected individuals. Immunogenicity in general refers to the potential of an antigen to elicit an immune response, while in case of allergens, allergenicity is considered as a reflection of its allergenic potential. Allergenicity indicates the capability of an allergen to induce clinical symptoms of allergy as well as to induce and bind to IgE antibodies [4]. The prevalence of allergic reactions has increased significantly in the last few years, especially in the developing countries [5]. This has resulted in considerable increase in disease burden as well as economic issues due to costs associated with these diseases. Therefore, the study of allergic diseases has gained tremendous importance as they represent one of the major health problems in urban and rural regions.

The field of allergy research has rapidly progressed in the last few years [6]. Recent advances in genomic, proteomic, and analytical methods have given rise to large amounts of data relevant to allergens. This data can be correlated with pathology of various allergic diseases based on experimental, clinical, and epidemiological data for allergic reactions. The continuous growth of data calls for the efficient archival, management, and analysis of data. This has led to the development of the field of allergen informatics, which comprises allergen specific databases/resources and computational methods/tools. Allergen informatics constitutes an important branch of immunoinformatics [7]. In this chapter, a review of the existing status of allergen informatics with respect to important aspects such as allergens and allergenicity,

allergen databases, algorithms/tools for allergen/allergenicity prediction, allergen epitope prediction, and allergenic cross-reactivity assessment has been presented (**Figure 1**).
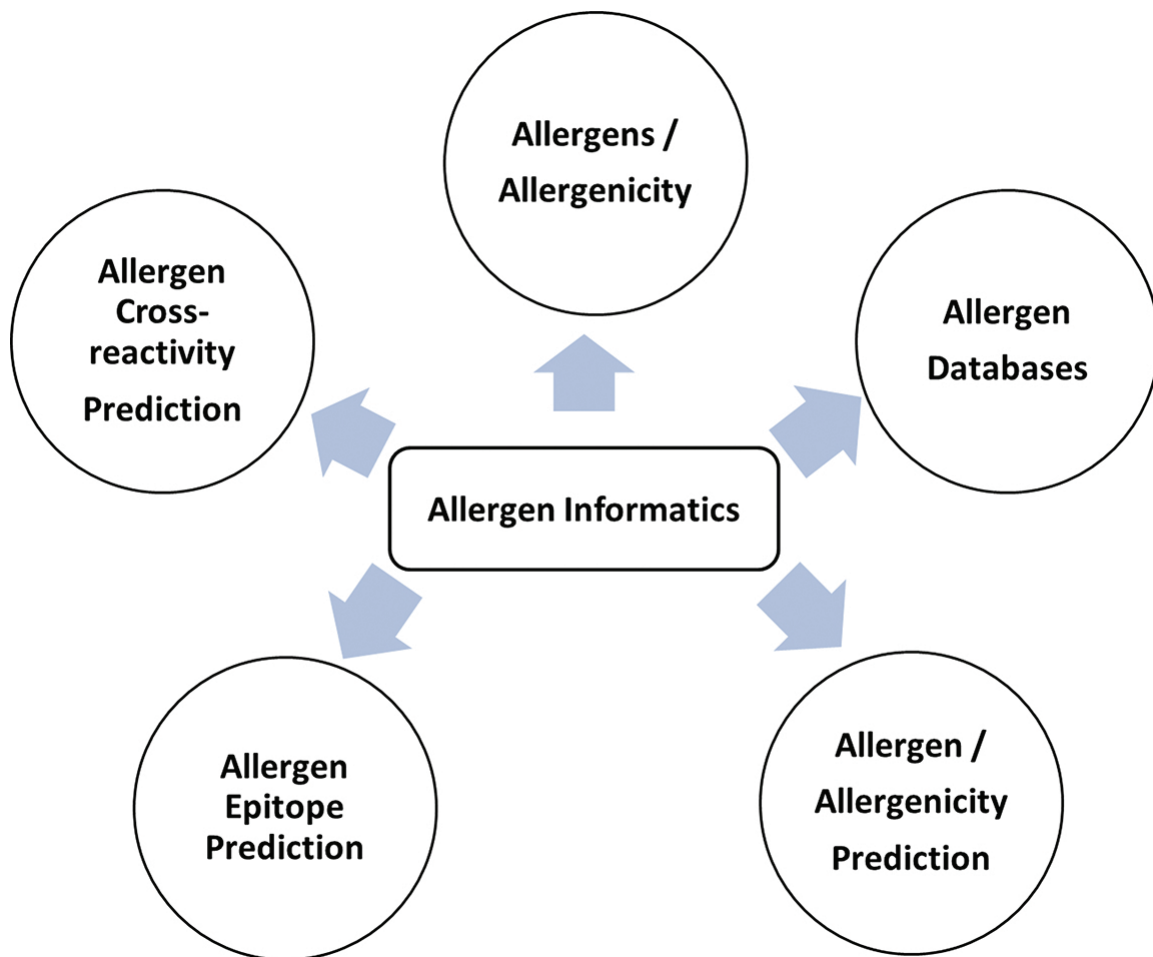

Figure 1. Topics covered for allergen informatics in the current review.

2. Allergens and allergenicity

Allergens represent the most critical component of an allergic reaction, although IgE antibody, Fc receptors, mast cells, and basophils as well as pharmacological mediators such as histamine and heparin also play very significant roles. Allergens are ubiquitous substances, which arise from a variety of sources such as foods, plants, animals, or environment. An allergen can either be a chemical substance (e.g., penicillin) or a protein (e.g., albumin, profilin, etc.). Majority of the allergens are proteins or glycoproteins that possess high water solubility. Several bio-chemical and structural features of allergens such as stability, hydrophobicity, and ligand-binding domains are known to contribute to their allergenicity [8]. However, common

molecular and structural features of allergens that are responsible for allergenicity have not yet been conclusively discovered.

Allergens are provided with a unique, unambiguous, and systematic nomenclature which has been developed and maintained by the World Health Organization (WHO) and International Union of Immunological Societies' (IUIS) "Allergen Nomenclature Sub-committee" [9, 10]. The nomenclature is based on the Linnean system, and an allergen, which satisfies certain biochemical and immunological criteria, is included in the WHO/IUIS nomenclature. An allergen name consists of an abbreviation of the scientific name of the allergen source organism. First 3–4 letters denote the genus name, while the subsequent 1–2 letters represent species, followed by an Arabic numeral that denotes the order of its identification. For instance, Der p 1 represents the first allergen to be characterized from the house dust mite *Dermatophagoides pteronyssinus*. An allergen may possess isoallergens or isoforms/variants, which are considered as multiple molecular forms of the same allergen. The WHO/IUIS nomenclature defines isoallergen as an allergen belonging to a single species, with a similar molecular size and identical biological function, and possessing ≥67% amino acid sequence identity while a variant or isoform corresponds to allergen sequences that differ by only a limited number of amino acid substitutions [11]. It is very important to archive and study the data on isoallergens and isoforms/variants in a differentiated manner as it has been shown that variations in allergens significantly affect their allergenicity and cross-reactivity as well as influence recognition of epitopes by T cells and IgE [12]. An allergen can be considered as a major or minor allergen based on the measure of its allergenicity. Major allergens are the ones to which >50% of patients with an allergy to its source are sensitized, while minor allergens are recognized by a limited number of patients [13].

Allergens display important features such as epitopes and cross-reactivity that are very critical with respect to understanding of allergic reactions and developing newer approaches for diagnosis and treatment of allergic diseases. Epitope or antigenic determinant refers to the immunologically active region of the allergen. An epitope can be an IgE-binding epitope or a T-cell epitope depending on whether it interacts with an IgE or a T-lymphocyte. An IgE epitope can be either sequential (linear) that consists of contiguous stretch of amino acids or conformational (discontinuous), which comprises amino acids present at different loci in an antigen. An antibody is said to be cross-reactive when it recognizes and binds to multiple antigens.

2.1. IgE-binding epitopes

IgE-binding epitopes refer to the IgE recognition sites in allergens that are involved in specific interaction of allergens and IgE antibody. Inferences drawn from allergen–antibody complexes and other important studies have shown that majority of IgE-binding epitopes are conformational in nature [14]. IgE epitopes possess some defining structural and immunological features such as they are more cross-reactive in nature and have higher intrinsic flexibility. These features make them distinct from other antibody epitopes and contribute significantly in the allergenicity [15, 16]. Identification and in-depth analysis of IgE-binding epitopes has the potential to contribute immensely in accurate diagnosis and allergen-specific immunotherapy of allergies, especially the food allergy [17, 18]. Large amount of data on allergen epitopes are

generated by employing strategies based on the use of overlapping synthetic peptides, recombinant allergenic fragments, cocrystal structure complexes, etc. However, it is believed that insights obtained from study of allergen–antibody complexes will be the most helpful in understanding the role these epitopes play in allergic reactions [19–21].

2.2. T-cell epitopes

T-cell epitopes are the antigenic determinants of allergens that interact with T-lymphocytes via specific T-cell receptors. T-cell epitopes of allergens have shown to be very important for the modulation of allergic response and thereby contributing to symptoms associated with allergic diseases [22]. They have enormous potential in the development of allergy vaccines as well as newer strategies in allergen immunotherapy, considering their fundamental role in allergic response [23, 24]. Recent findings have indicated that T-cell epitope repertoire in allergens is diverse than IgE epitopes, and it can be very useful in specific immunotherapy in allergy [25]. An analysis carried out on available epitope data has shown that T-cell epitopes are known to occur more commonly in the airborne allergens as compared to food allergens [26].

2.3. Cross-reactivity

Cross-reactivity denotes a clinically and immunologically critical phenomenon displayed by allergens from various sources and is the cause of pollen-food syndromes, such as the one seen in case of birch and apple. Cross-reactivity is considered as a property of antibodies and it arises when an antibody or a subgroup of antibodies recognizes more than one allergen or epitope [27]. Two allergens are considered cross-reactive if they are recognized by a single antibody (or T-cell receptor). It has been stated that cross-reactivity among allergens at the level of B cells, T cells, and mast cells reflects clinical sensitivities and contributes very significantly in the regulation of allergic sensitization [28].

Cross-reactivity is predominantly an antibody defined phenomenon and IgE antibodies are shown to be more cross-reactive in nature. Affinity of the antibodies toward the allergen is known to play an important role in cross-reactivity. However, the properties of the allergenic protein are also very important and shared features on the level of both primary and tertiary structures of the cross-reactive proteins are found to be responsible for cross-reactivity [4]. Similarity at the level of sequence is an important indicator and cross-reactivity seems to require more than 70% sequence identity. In addition to this, other factors such as the host immune response against the allergen, dosage of allergen, and mode of exposure also contribute in clinical relevance of allergic cross-reactivity. Inferences drawn from studying a large number of allergens have led to the conclusion that structural similarity among proteins from diverse sources is the molecular basis of allergic cross-reactivity [29]. Considering the role it plays in the development of allergic symptoms, a detailed analysis of cross-reactivity has the potential to contribute in the development of new strategies in diagnosis and therapy of allergic diseases.

3. Allergen databases

Last few years have witnessed substantial technological advances in the field of genomics and proteomics along with tremendous improvements in analytical methods. This has led to a significant progress in the area of allergy research. As a result of this, there has been a steady and continuous increase in the number of characterized protein allergens over the last few years. Efficient storage and management of data has become very important because of such incessant accumulation of molecular and clinical data on allergens. Therefore, allergen databases represent very crucial resources for basic allergy research as they are involved in archival of available allergen knowledge.

Database	Developed by (URL)	Type of data archived	Computational tools (if any)	Updates
IUIS Allergen [36]	WHO/IUIS Allergen Nomenclature Sub-committee (http://www.allergen.org)	Sequence (isoallergens/ isoforms), structure, allergenicity	–	Updated continuously
Allergome [39]	Centre for Clinical and Experimental Allergology, Italy (http://www.allergome.org)	Sequence (isoallergens/ isoforms), structure, clinical, epidemiological, cross-reactivity, etc.	–	Updated continuously
Structural Database of Allergenic Proteins (SDAP) [41]	Sealy Centre for Structural Biology, University of Texas, USA (https://fermi.utmb.edu)	Sequence, structure, structural models, IgE epitopes	Yes	2013
Allergen Database For Food Safety (ADFS) [44]	National Institute of Health, Japan (http://allergen.nihs.go.jp/ADFS/)	Sequence, structure, IgE epitopes, small molecule allergens	Yes	2016
AllergenOnline [46]	Food Allergy and Resource Program (FARP) (http://www.allergenonline.org)	Sequence, allergenicity	Yes	2016
AllFam [48]	Department of Pathophysiology and Allergy Research, Medical University of Vienna, Austria (http://www.meduniwien.ac.at/allfam)	Allergen family data, cross-link to Pfam database	–	2011
AllergenPro [53]	The National Agricultural	Sequence, IgE epitopes	Yes	2015

Database	Developed by (URL)	Type of data archived	Computational tools (if any)	Updates
	Biotechnology Information Centre, Korea (http://nabic.rda.go.kr/ allergen)			
AllerBase [61]	Bioinformatics Centre, Savitribai Phule Pune University, India (www.bioinfo.net.in/ AllerBase/Home.html)	Sequence and structure (cross-links), IgE epitopes, IgE antibody, IgE cross-reactivity, experimental evidences of allergenicity	–	Updated continuously

Table 1. Summary of major allergen databases and their relevant features.

Many allergen-specific databases have been developed in the past few years although they differ from each other with respect to their objectives, type of data archived, accessibility of contents, and the level of annotation and applications [30]. In addition to dedicated allergen databases, primary bioinformatics databases also document significant data on allergens. Examples of these databases include GenBank/GenPept [31, 32], UniProtKB [33], and Protein Data Bank (PDB) [34], which archive sequence and structure data on allergens along with its annotation. The Summary of allergen-specific databases is provided in **Table 1**. In the following section, the existing allergen databases are described.

3.1. IUIS Allergen database

The IUIS Allergen Nomenclature Sub-Committee, under the auspices of the WHO, provides the systematic nomenclature of allergenic proteins and it has developed and maintained Allergen database [35, 36]. The database archives all of the WHO/IUIS–recognized allergens along with their isoallergens and isoforms (variants). In order to maintain a consistent allergen nomenclature for newly discovered allergens, researchers are required to submit newly described allergens to the Allergen Nomenclature Sub-Committee before submitting their manuscript to a journal for consideration for publication.

Each allergen in this database is provided with annotation that includes biochemical name, molecular weight, information on its allergenicity, reference, etc. Additionally, sequence data for allergens and isoallergens/isoforms are also stored in the database, along with cross-references to GenBank [31], GenPept [32], and UniProtKB [33], as well as to PDB [34], for nucleotide, protein sequences, and 3D structure data, respectively. Allergen database can be searched by using allergen name, biochemical name, allergen source organism, taxonomic group, etc., as search criteria. The database is updated continuously with specific names assigned to newly discovered allergens and isoallergens/variants [37]. Allergen database does not exemplify the comprehensive allergen data although it documents majority of the characterized allergens. This is because there are a large number of allergens that have been

reported in literature which are not recognized by IUIS-Allergen. The database does not archive data on allergen epitopes and cross-reactivity.

3.2. Allergome database

Allergome [38] represents an extensive repository of information on allergen molecules causing IgE-mediated (allergic, atopic) diseases [39]. The database comprises comprehensive data on WHO/IUIS-approved allergens along with other non-recognized allergens. These allergenic molecules are selected and curated from the published literature and web-based resources. It also contains data on allergenic sources based on whether they possess identified molecules or not. Allergome documents information on allergen and isoallergens/isoforms along with their sequences. Cross-links to sequence and structure databases like UniProtKB [33] and PDB [34] are also provided.

Allergome can be searched by using basic and advanced search options. Basic search employs numerous search criterions such as allergen name, biochemical name, source organism, etc., while advanced search enables the user to search using specific attributes. Each allergen molecule is represented by a monograph which represents information about the three parts of allergen such as basic information, data on the native form, and its recombinant form. The most important and unique feature of Allergome platform is the presence of several support modules that deal with archival of specific aspects of allergen data. A couple of important modules are RefArray, for easy access to references stored in the Allergome, and Real Time Monitoring of IgE sensitization (ReTiME), for real-time data collection and storage of IgE sensitization data and the number of other utilities. Allergome is updated regularly and allergen data curated from literature is documented.

3.3. Structural database of allergenic proteins

Structural Database of Allergenic Proteins (SDAP) [40] is an allergen database that prominently deals with structural aspects of allergens [41]. It houses comprehensive cross-referenced sequence data on allergens, IgE-binding epitopes, 3D structures, and models of allergens. Each allergen in SDAP is provided with cross-links to primary databases such as UniProtKB [33], PDB [34], as well as to important resources such as NCBI Taxonomy Browser [42] and PubMed [42] for literature references. SDAP also has a utility as a web server that integrates various computational tools, which assist structural biology–related studies dealing with allergens and their epitopes. It employs an algorithm based on the conserved properties of amino acid side chains to detect regions associated with allergenicity in novel sequences. The database consists of number of tools that can be used to assess potential cross-reactivity of allergens and also help in screening of IgE epitopes in novel proteins. The last update of the database was carried out on February 25, 2013. SDAP does not archive complete data for allergens that are not recognized by IUIS while data on allergen cross-reactivity is also not documented.

3.4. Allergen database for food safety

Allergen Database for Food Safety (ADFS) [43] is developed as a project of the Division of Biochemistry and Immunochemistry of National Institute of Health Sciences (Japan). The aim of the database is to archive allergenic proteins and their IgE epitopes with a special emphasis on food allergens and food safety [44]. Allergens archived in ADFS are grouped into eight categories such as pollen, mite, animal, fungus, insect, food, latex, and others, and each allergen entry is provided with the primary database accession numbers of their genes and 3D structure information. The database is also equipped with homology-based sequence search tool for the evaluation of allergenicity. One of the most distinct features of ADFS is the archival of data on small molecule, nonprotein (chemical) allergens. The database does not archive data on allergen cross-reactivity.

3.5. AllergenOnline database

AllergenOnline [45] is a well curated allergen database that documents a peer reviewed allergen list, which is compiled from various resources such IUIS-Allergen, PubMed, scientific publications, and other allergen databases. The database was developed within the Food Allergy Research and Resource Program (FARRP) at the University of Nebraska [46]. For each allergen, data on source organism, common name, IUIS official nomenclature, protein length, class of allergen like food allergen, contact allergen, etc., and a link to the NCBI protein (GenPept) [32] database are provided. AllergenOnline also provides the utility for sequence-based searches for allergens, which include alignments by FASTA and an eight-amino acid short-sequence identity search. This utility can be very useful in the identification of proteins that may present a potential risk of allergenic cross-reactivity. AllergenOnline is updated every year and the last update that resulted in version 16 of the database was reported on January 27, 2016. It does not archive data on allergen epitopes as well as on allergenic cross-reactivity.

3.6. AllFam database

AllFam [47] represents a very important resource for allergens as it is involved in classification of allergens into protein families [48]. This study has shown that allergens are distributed into relatively few protein families and possess a limited number of biochemical functions. The structural classification of allergens in AllFam is performed by using family information from PFam [49] and the Structural Classification of Proteins (SCOP) database [50], while biochemical functions of allergens were extracted from the Gene Ontology annotation database [51]. The database provides the option of browsing lists of allergen families based on allergen source (plants, animals, and fungi) and route of exposure (inhalation, ingestion, etc.) while search for specific protein families can also be performed. Each allergen family in AllFam is linked to a family fact sheet that describes the biochemical properties of the family members as well as a list of key references related to this family. The last update of AllFam was reported on September 12, 2011. AllFam does not archive data on molecular features of individual allergens although cross-link to IUIS-Allergen and Allergome is provided for each documented allergen.

3.7. AllergenPro database

AllergenPro [52] is a recently developed allergen database that archives data on allergen sequences, structures, and epitopes from various sources. It is an integrated database which provides information about allergens in foods, microorganisms, fungi, animals, and plants [53]. It has been provided with a utility to search for allergens based on keywords as well as the sequence. AllergenPro is also equipped with a computational tool for the prediction of allergenicity. Prediction is based on three different approaches such as FAO/WHO guidelines (sequence)–based approach, motif-based approach, and epitope-based approach. The database was last updated on June 4, 2015. AllergenPro does not archive data on allergen cross-reactivity while the literature references for documented allergens and epitopes have also not been provided.

3.8. Archival of allergen epitope data

As mentioned earlier, epitopes denote very important feature of allergens as they play vital role in allergic diseases. Because the molecular characterization of allergens has risen immensely in recent years, the data on allergenic epitopes has also increased significantly. Therefore, it has become necessary to store and manage the epitope data for its efficient utilization.

Some of the existing allergen databases described above, such as SDAP, ADFS, and AllergenPro are involved in storage of allergen epitope data. There are few databases available that are dedicated for epitope data from all types of antigens, which also document information on allergen epitopes [54–56]. However, the allergy-associated epitope data stored in these databases may not be comprehensive. The Immune Epitope Database (IEDB) [57], which is a repository of immune epitope reactivity data, is also a major database of allergy-derived epitope data [58]. It archives extensive allergen epitope data along with biological assays associated with them, including IgE-binding as well as T-cell epitopes curated and compiled from allergy-related references. IEDB is also equipped with several strategies for efficient searching and visualization of data on allergy-related epitopes [59]. Therefore, it represents a very useful and user friendly platform to access and retrieve allergy-related epitope data for the community of allergists. In a study involving classification of all the epitope-specific literature in various immunological domains, it is stated that IEDB comprises relatively fewer references for allergy-derived epitopes as compared to Cancer and Infectious Diseases [60]. This indicates that there is considerable scope for more in-depth archival of allergen epitope data from literature. Another study on meta-analysis of the allergy-associated epitope data in IEDB has indicated that relatively lesser data is archived for allergen T-cell epitopes as compared to IgE epitopes [26].

3.8.1. AllerBase database

Observations from the study of the existing allergen databases indicated that they archive significant data on various aspects of allergen and allergenicity, although the level of completeness of data differs considerably for diverse allergen features. AllerBase [61] is a recently

developed comprehensive database of allergens and allergen features which addresses some of the limitations associated with the existing allergen databases [Kadam et al. 2016, unpublished]. The database comprises extensive data on experimentally validated allergens and allergen specific features such as IgE-binding epitopes, IgE cross-reactivity, IgE antibodies, and evidences for experimental validation of allergens. AllerBase is provided with basic and advanced search utilities along with browse database option to retrieve desired allergen data. The Completeness Index, which represents availability of data for various features for each allergen and a structure visualization utility, denote important features of the database. AllerBase also provides cross-references to several immunological and allergen databases and represents a notable instance of integration of allergen data from number of resources.

4. Computational prediction of allergens/allergenicity

Allergens mainly comprise commonly occurring proteins in foods, pollens, and other biological entities in the environment. It has become necessary to assess the potential allergenicity of these proteins considering the health hazards associated with allergic reactions to them. In recent years, genetic engineering and food processing methods are routinely employed for modifying the existing proteins or introducing new ones. Analysis of allergenicity of such proteins/products along with newly introduced biopharmaceuticals is absolutely essential in order to avoid transfer of an allergenic molecule. Computational assessment or prediction of allergenicity represents the major approach to test for allergenicity, and numerous bioinformatics tools/methods have been employed successfully for this purpose [62]. The majority of these methods utilize the amino acid sequence of allergens along with its different features, while a very few approaches use structure information. **Table 2** denotes the list of computational tools/servers available for the prediction of allergens/allergenicity. In the following section, the prominent approaches used for the computational assessment or prediction of allergens/allergenicity are described briefly.

No.	Method (URL)	Approach used	Efficiency
1	SDAP (http://fermi.utmb.edu/) [41]	FAO/WHO guidelines	–
2	AllergenOnline (http://www.allergenonline.org) [46]	FAO/WHO guidelines	–
3	AllergenPro (http://nabic.rda.go.kr/allergen) [53]	FAO/WHO guidelines, sequence motifs, epitopes	–
4	Allermatch (http://allermatch.org) [66]	FAO/WHO guidelines	–
5	AllerTool (http://research.i2r.a-star.edu.sg/AllerTool/)	FAO/WHO guidelines, global representation	$A_{ROC} = 0.90$, SE = 86%, SP = 86%

No.	Method (URL)	Approach used	Efficiency
	[67]	of protein sequence and SVM	
6	WebAllergen (http://weballergen.bii.a-star.edu.sg/) [73]	Sequence motifs	–
7	AlgPred (http://www.imtech.res.in/raghava/algpred/) [75]	Sequence features and SVM, sequence motifs, epitopes, allergen representative peptides	Accuracy = 85%, SE = 88%, SP = 81%
8	AllerTOP (http://www.ddg-pharmfac.net/AllerTOP) [83]	Sequence based descriptors, auto and cross-covariance, machine learning	Accuracy = 85.3%, SE = 82.5%, SP = 88.1%
9	EVALLER (http://bioinformatics.bmc.uu.se/evaller.html) [86]	DFLAP algorithm and SVM	–
10	AllerHunter (http://tiger.dbs.nus.edu.sg/AllerHunter/) [89]	Iterative pairwise sequence similarity and SVM	A_{ROC} = 0.928, accuracy = 95.3%, SE = 83.4%, SP = 96.4%
11	APPEL (http://jing.cz3.nus.edu.sg/cgi-bin/APPEL) [91]	Sequence based features and SVM	MCC = 0.95, SE = 93%, SP = 99.9%
12	SORTALLER (http://sortaller.gzhmu.edu.cn/) [93]	AFFP dataset, normalized BLAST E-values and SVM	MCC = 0.97, SE = 98.6%, SP = 98.4%
13	PREAL (http://gmobl.sjtu.edu.cn/PREAL/index.php) [96]	Biochemical and physicochemical descriptors, sequence features, subcellular locations, mRMR, SVM	Accuracy = 93.42%
14	Allerdictor (http://allerdictor.vbi.vt.edu/) [99]	Sequences as text documents, Naive Bayes classifier and SVM	–
15	proAP (http://gmobl.sjtu.edu.cn/proAP/main.html) [100]	Integration of methods based on FAO/WHO guidelines, sequence motifs and SVM	–
16	AllergenFP (http://ddg-pharmfac.net/AllergenFP/) [103]	Auto and cross-covariance, descriptor-based fingerprints	Accuracy = 88%, MCC = 0.759

No.	Method (URL)	Approach used	Efficiency
		of residues	
17	FuzzyApp (http://fuzzyapp.bicpu.edu.in/fuzzyapp.php) [107]	Fuzzy rule based system	–

A_{ROC}: area under the receiver-operating curve; SE: sensitivity; SP: specificity; MCC: Matthews correlation coefficient.

Table 2. List of computational tools/servers for allergen/allergenicity prediction.

4.1. Sequence similarity-based approaches

One of the first studies dealing with analysis of allergenicity was put forth by Metcalfe et al. [63]. They have proposed a decision tree–based approach for allergenicity assessment of foods derived from genetically modified crops. The first computational approach for the assessment of allergenicity was provided by "Codex Alimentarius Commission" of FAO/WHO [64, 65]. It stated that a protein can be regarded as an allergen if it consists of an exact match with at least six contiguous amino acids or showed more than 35% similarity over a window of 80 amino acids when compared with a sequence of known allergen. This approach has been widely used to predict allergenicity and there are number of web servers for allergen prediction, which are based on it. Allermatch [66], AllerTool [67], and AllergenPro [53] are some of the prominent web servers which employ these FAO/WHO guidelines for allergen prediction. Additionally, some of the major allergen databases such as SDAP [41] and AllergenOnline [46] also utilize this strategy for allergenicity prediction. A recent study performed by Verma et al. [68] has shown that the sequence similarity-based approach gives substantially better results when used in combination with other bioinformatics methods. However, results obtained by certain studies indicated that approaches based on these guidelines are not highly efficient for identifying allergenic proteins and many of times they lead to false or irrelevant allergenicity estimations [69–71]. As a result of these observations, it became necessary to discover and employ other strategies for the prediction of allergenicity.

4.2. Motif-based approaches

In a study carried out by Stadler and Stadler [71], it was observed that the use of sequence motifs, which represent the secondary structures of proteins, performs significantly better than the approach based on FAO/WHO guidelines. This method employs MEME motifs of a length of 50 residues for the prediction of allergenicity by using pairwise sequence alignment with certain threshold. WebAllergen [72] is a web server for the prediction of allergenic proteins which is also based on specific detectable allergenic motifs in known allergens [73]. Furthermore, a study carried out by Kong et al. showed that an approach based on search of multiple motifs is more specific and efficient than the conventional single motif search [74]. AlgPred [75] and AllergenPro [53] are important web servers for allergen prediction in which one of the prediction approaches is based on allergen-derived motifs. A recent study that employs computational approaches for comparison of allergens and metazoan parasite proteins stated that significant sequence and structure similarity exists between parasite proteins and

allergenic proteins [76]. The analysis was carried out using sequence and structural motifs in allergens and a workflow was developed for the computational analysis of parasite proteins.

4.3. Machine learning–based approaches

Recent years have witnessed tremendous increase in the application of machine learning methods for solving biological problems. Machine learning–based approaches have been widely used for predicting various aspects of protein function [77]. These methods are also employed routinely for the development of algorithms to predict allergenicity of novel proteins.

Although Support Vector Machine (SVM) is the most commonly used machine learning method for allergen prediction, other methods have also been frequently employed. One of the earliest methods was developed by Zorzet et al. [78] that utilizes a k-Nearest-Neighbor (kNN) classification algorithm for the prediction of allergenicity, while a Bayesian classifier was employed by Soeria-Atmadja et al. [79] for the same purpose. An approach based on the combination of hidden Markov model (HMM) and conserved motifs in allergen was also used to successfully predict protein allergenicity [80]. Dimitrov et al. [81] developed two artificial neural network (ANN)-based algorithms for allergenicity prediction, which utilize descriptors derived from amino acids that denote their structural and physicochemical properties. AllerTOP [82] is an online bioinformatics tool to perform the computational prediction of allergens [83]. This algorithm employs descriptors that denote the chemical properties of amino acids in allergen sequences and auto- and cross-covariance transformation along with five machine learning methods for classification. These methods are random forest, multilayer perceptron, logistic regression, decision tree, naïve Bayes, and kNN.

There are number of web-based tools/servers developed which use SVM for performing classification/prediction of allergens. AlgPred [84] is one of the earlier web servers developed for the prediction of allergenic proteins [75]. It employs SVM with amino acid and dipeptide composition as features of allergens to achieve accuracy of 85.02 and 84.00%, respectively. EVALLER [85] is another web server created for *in silico* determination of potential allergenicity with very good efficiency [86]. It performs detection based on filtered length–adjusted allergen peptides (DFLAP) algorithm and SVM. AllerTool [87] web server also applies SVM-based algorithm for the prediction of allergenicity and provides sensitivity and specificity of 86.00% [67]. AllerHunter [88] is an important web-based computational system for allergenicity assessment which uses a scheme based on iterative pairwise sequence similarity encoding along with SVM [89]. The method is very efficient with a sensitivity of 83.4% and a specificity of 96.4%.

A web-based tool APPEL [90] is developed for the prediction of allergenic proteins that employs physicochemical and structural features derived from allergen sequence in combination with SVM [91]. Zhang et al. have developed an online allergen prediction tool titled SORTALLER [92], which is based on allergen family featured peptide (AFFP) dataset and employs SVM as a classifier [93]. An algorithm developed by Mohabatkar et al. [94] for the prediction of allergenic proteins utilizes pseudoamino acid composition (PseAAC) along with SVM and provides an accuracy of 91.19%. PREAL [95] is web-based tool that performs allergen

prediction by using SVM along with feature selection methods such as maximum relevance minimum redundancy (mRMR) and incremental feature selection (IFS) [96]. A combination of hydrophobicity amino acid index and discrete Fourier transform along with an SVM classifier is employed for highly efficient prediction of allergenicity in a signal-processing bioinformatics approach [97]. Allerdictor [98] is web server that specializes in large-scale allergen discovery. It models protein sequences as text documents and employs SVM in text classification for carrying out allergen prediction [99].

4.4. Other approaches

A study carried out by Wang et al. evaluated sequence-, motif-, and SVM-based approaches for the computational prediction of allergens and also performed parameter optimization to obtain better performance [100]. The resulting methods from this study are integrated and made available as a web application titled proAP [101]. AllergenFP [102] is a recently developed web server for allergenicity prediction that utilizes alignment-free descriptor-based fingerprint approach [103]. The descriptors used here are important properties of amino acid such as size, hydrophobicity, relative abundance, helix, and beta-strand forming propensities, etc. In a structure-based approach proposed by Bragin et al. [104], information derived from protein 3D structure is used for the representation of protein surface as patches designated as discontinuous peptides. It is observed that prediction of allergenic proteins based on this approach gave better accuracy. Vijayakumar and Lakshmi have developed a fuzzy inference system–based algorithm for allergenicity prediction that utilizes five different modules [105]. These modules consist of a machine learning classifier, motif search, sequence similarity, FAO/WHO evaluation scheme, etc. FuzzyApp [106], a web server based on fuzzy rule–based system, is then developed for the prediction of allergenicity [107]. Jiang et al. performed an analysis of food allergens using a computational model that simulates gastric fluid digestion [108]. This study stated that food allergens could be classified as alimentary canal-sensitized and nonalimentary canal-sensitized allergens based on the digestibility of these allergens in simulated gastric fluid.

5. Computational prediction of allergen epitopes

Epitopes represent distinctive amino acid residues on the antigens and are important determinants of an immune response. Identification of epitopes is considered a key aspect of designing highly effective multiple-subunit vaccines and developing efficient diagnostic and therapy methods against allergens. Although experimental methods have been very useful for the identification of epitopes, their usefulness is restricted because of their time- and cost-intensive nature and inability in dealing with large-scale elucidation of epitopes. Hence, computational approaches are considered to be very beneficial alternative as they are cost and time effective.

Large number of highly efficient algorithms and tools have been developed over the years for the computational prediction of epitopes. These methods deal with the prediction of

both B-cell and T-cell epitopes as well as sequential (linear) and discontinuous (conformational) epitopes. Based on the information (data) utilized for performing prediction, the methodologies can be grouped as sequence-based or structure-based approaches. Many sequence-based linear epitope prediction methods for B cells have been developed and used since long time and majority of them are propensity scale and machine learning–based methods [109, 110]. Some of the major tools/servers that deal with the prediction of linear B-cell epitopes are listed in **Table 3**.

No.	Method (URL)	Approach used	Efficiency
1	ABCPred (http://www.imtech.res.in/raghava/abcpred/) [111]	Fixed length epitope patterns, recurrent ANNs	Accuracy = 65.93%, SE = 67.14%, SP = 64.71%
2	APCpred (http://ccb.bmi.ac.cn/APCpred/) [112]	Amino acid anchoring pair composition (APC) and SVM	A_{ROC} = 0.809, accuracy = 72.94%
3	BCPreds (http://ailab.cs.iastate.edu/bcpreds/) [113]	SVM classifiers with string kernels	A_{ROC} = 0.758
4	BcePred (http://www.imtech.res.in/raghava/bcepred/) [114]	Physicochemical properties of epitope residues	Accuracy = 58.7%
5	BepiPred (http://www.cbs.dtu.dk/services/BepiPred/) [115]	Parker's hydrophilicity scale and HMM	–
6	BEST (http://biomine.ece.ualberta.ca/BEST/) [116]	Antigen sequence features, SVM	A_{ROC} = 0.85
7	Bayesb (http://www.immunopred.org/bayesb/index.html) [117]	Bayes feature extraction and SVM	A_{ROC} = 0.84, accuracy = 74.5%
8	COBEpro (http://scratch.proteomics.ics.uci.edu) [118]	Antigen fragment score and SVM	A_{ROC} = 0.829
9	EPMLR (http://www.bioinfo.tsinghua.edu.cn/epitope/EPMLR/) [119]	Sequence features and multiple linear regression (MLR)	A_{ROC} = 0.728, SE = 81.8%, SP = 64.1%
10	IEDB Analysis Resource (http://tools.iedb.org/bcell/) [120]	A collection of tools based on various methods	–
11	LBtope (http://www.imtech.res.in/raghava/lbtope/) [121]	Large datasets of epitopes, KNN, SVM	Accuracy = 86%
12	SVMTriP (http://sysbio.unl.edu/SVMTriP/) [122]	Tri-peptide similarity, propensity scores and SVM	A_{ROC} = 0.702, SE = 80.1%, SP = 55.2%

A_{ROC}: area under the receiver-operating curve; SE: sensitivity; SP: specificity; MCC: Matthews correlation coefficient.

Table 3. List of major tools/servers for sequential (linear) epitope prediction.

Number of methods that utilize 3D structure of antigens for discontinuous epitope prediction have also been developed. These methods use different approaches for prediction such as

solvent accessibility of surface residues [123, 124], solvent accessibility with propensity scores [125], and propensity scores with packing density of amino acids [126]. An account of major tools/servers that are involved in conformational epitope prediction is provided in **Table 4**.

No.	Method (URL)	Approach used	Efficiency
1	BEpro (formerly PEPITO) (http://pepito.proteomics.ics.uci.edu/) [127]	3D structure of antigen, amino acid propensity scores	$A_{ROC} = 0.75$
2	B-Pred (http://immuno.bio.uniroma2.it/bpred) [128]	3D structure or model of antigen, solvent exposure of residues	$SE = 0.70$
3	CBTOPE (http://www.imtech.res.in/raghava/cbtope/) [129]	Sequence features and SVM	$A_{ROC} = 0.9$, Accuracy $= 85\%$
4	CEP (http://196.1.114.49/cgi-bin/cep.pl) [124]	3D structure of antigen, solvent accessibility of amino acids	Accuracy = 75%
5	DiscoTope 2.0 (http://www.cbs.dtu.dk/services/DiscoTope/) [125]	3D structure of antigen, epitope propensity scores, surface accessibility	$A_{ROC} = 0.824$
6	ElliPro (http://tools.immuneepitope.org/tools/ElliPro) [130]	3D structure of antigen, Thornton's method, residue clustering algorithm	$A_{ROC} = 0.732$
7	Epitopia (http://epitopia.tau.ac.il/) [131]	Antigen sequence or 3D structure, Naïve Bayes classifier	$A_{ROC} = 0.59$
8	EPSVR (http://sysbio.unl.edu/EPSVR/) [132]	3D structure of antigen, Support vector regression (SVR)	$A_{ROC} = 0.597$
9	EPMeta (http://sysbio.unl.edu/EPMeta/) [132]	Meta server integrating EPSVR with other methods	$A_{ROC} = 0.638$
10	EPCES (http://sysbio.unl.edu/EPCES/) [133]	3D structure of antigen, surface features	–
11	SEPPA 2.0 (http://badd.tongji.edu.cn/seppa/) [126]	3D structure of antigen, subcellular localization of antigen, residue propensity, etc.	$A_{ROC} = 0.745$

A_{ROC}: area under the receiver-operating curve; SE: sensitivity; SP: specificity; MCC: Matthews correlation coefficient.

Table 4. List of major tools/servers for conformational (discontinuous) epitope prediction.

Studies have shown that the analysis of antigen–antibody complex structures is very useful for the characterization of conformational epitopes [134]. A dedicated resource titled AgAbDb [135] that archives the interactions derived from antigen–antibody complexes is available, which can be very useful for the analysis of epitopes [20, 21]. Several algorithms have also been developed for the prediction of T-cell epitopes in antigens. These methodologies deal with the prediction of peptides that possess the ability to interact with specific major histocompatibility complex (MHC) molecules [136]. Machine learning–based approaches are very commonly employed for this purpose and are found to be very efficient [137]. The details of epitope prediction methods/tools for B cells and T cells have been reviewed elsewhere [138, 139]. Some

of the important tools/servers that perform the prediction of T-cell epitopes are listed in **Table 5**. Recently, it has been shown that epitope prediction can be performed over the whole proteome by integrating multiple epitope prediction methods [149]. Antibody-specific epitope prediction has emerged as a significant alternative to the traditional antibody-independent epitope prediction methods [150].

No. Method (URL)	Approach used	Efficiency
1 CTLPred (http://www.imtech.res. in/raghava/ctlpred/index.html) [140]	Cytotoxic T-lymphocyte epitopes, SVM, ANN	Accuracy = 75.2%
2 EpiJen (http://www.ddg-pharmfac. net/epijen/EpiJen/EpiJen.htm) [141]	Multi-step algorithm that employs integrated approach	Accuracy = 60%
3 EpiTOP (http://www.pharmfac. net/EpiTOP/) [142]	QSAR approach based on proteochemometrics	Accuracy = 89%
4 NetCTLpan (http://www.cbs.dtu. dk/services/NetCTLpan/) [143]	Integrated method employing proteasomal cleavage, TAP transport efficiency, and MHC class I binding affinity	A_{ROC} = 0.95
5 NetMHCIIpan-3.0 (http://www.cbs.dtu.dk/ services/NetMHCIIpan-3.0/) [144]	A method for all HLA class II molecules based on peptide-binding MHC environment	A_{ROC} = 0.807
6 PREDIVAC (http://predivac. biosci.uq.edu.au/) [145]	Based on specificity-determining residues	–
7 SYFPEITHI (http://www.syfpeithi.de/bin/ MHCServer.dll/EpitopePrediction.htm) [146]	Scoring system based on position of residue in the epitopes	–
8 TEPITOPEpan (http://www.biokdd.fudan. edu.cn/Service/TEPITOPEpan/) [147]	Algorithm based on HLA-DR binding pocket similarity	–
9 WAPP (http://abi.inf.uni-tuebingen.de/ Services/WAPP/) [148]	Combination of methods based on proteasomal cleavage, TAP transport and MHC binding	–

A_{ROC}: area under the receiver-operating curve.

Table 5. List of major tools/servers for T-cell epitope prediction.

Epitopes represent critical components of allergens from the perspective of allergic reactions and development of new diagnosis and treatment strategies. Therefore, the computational prediction of these epitopes in allergens is of immense importance. Due to limitations associated with detailed archival of allergen epitope data and highly heterogeneous nature of the data, the number of tools available for allergen epitope prediction are far less, especially when compared with the number of tools available for allergen/allergenicity prediction. Therefore, the general epitope prediction methods listed above can be employed for epitope prediction

studies in allergens. Kleter and Peijnenburg [70] developed a strategy to screen for the potential linear IgE-epitopes using sequence comparison with a minimal length of six amino acids. The approach was moderately effective and it showed that further verification of IgE binding of epitopes by experimental tests is necessary. AlgPred [84] developed by Saha and Raghava is one of the first and major tools for the computational assessment of IgE epitopes [75]. Here, a database of known IgE epitopes is created and it is used to accurately predict allergenic proteins. AllerPred is a SVM-based computational system for the assessment of overlapping continuous and discontinuous B-cell epitope binding patterns in allergenic proteins [151]. This approach is successfully used to predict allergenicity of novel proteins. Dall'Antonia et al. [152] have developed a software tool titled Surface comparison–based Prediction of Allergenic Discontinuous Epitopes (SPADE). The algorithm consists of a structure-based comparison of allergen surfaces and IgE cross-reactivity data and is able to predict IgE epitopes from three important allergen families. A recent work performed by Lollier et al. [153] on meta-analysis of IgE-binding epitopes provided some important findings regarding these epitopes. They computed the fraction of allergen amino acids that are involved in epitopes and modeled a relationship between the rising number of literature references and the amino acid fractions to assess the possibility of binary classification of epitopes and nonepitopes. A web-based tool LocAllEpi [154] is also developed for the visualization of allergen epitopes along the protein sequence and their structural features.

6. Computational prediction of allergenic cross-reactivity

Cross-reactivity plays an important role in allergic reaction from the immunological and clinical context. Therefore, the computational prediction of allergenic cross-reactivity has been considered of substantial significance. The prediction of cross-reactivity in allergens is associated with the prediction of allergenicity for the majority of the cases. This is mainly because the antigenic determinants that contribute to the cross-reactivity in allergens are also responsible for their allergenicity. As a result of this, many of the tools/algorithms that have been developed for the prediction of allergens/allergenicity also perform cross-reactivity prediction.

The criteria defined by FAO/WHO experts, which have been mentioned earlier, help to identify cross-reactivity in allergens [155]. AllerTool [87] is a web server that performs cross-reactivity prediction based on amino acid sequence and WHO/FAO guidelines [67]. It also provides a graphical representation of the published and predicted cross-reactivity patterns of allergens. Stadler and Stadler [71] developed a sequence-based approach and stated that motif-based strategy provides better results for the computational assessment of cross-reactivity than the FAO/WHO guidelines. SDAP [40], which is a specialized allergen database described before, also comprises a sequence-based tool for the identification of cross-reactivity among allergens [41]. AllerHunter [88] is a SVM-based web server that deals with efficient assessment of allergic cross-reactivity in proteins [89]. A recently developed fuzzy inference system–based algorithm for allergenicity prediction is also able to predict cross-reactivity in allergens [105].

7. Future perspectives and challenges

Allergy represents a serious problem, as allergic diseases are known to affect millions of people worldwide. Advancements in genomic, proteomic, and analytical techniques have led to the generation of large amount of data related to allergy and allergens. Archival and analysis of these data denotes a major challenge in allergen bioinformatics. Data integration is one of the key limitations for efficient and useful storage of allergy associated data. This is mainly due to the heterogeneous nature of the data, which is derived from various sources such as molecular data from experimental characterization of allergic reactions, clinical, and epide-miological data from patients/populations. Bioinformatics resources and tools have an important role to play in overcoming this problem. In the wake of ever-expanding volume of data, it is vital to focus on developing databases/resources that will integrate information from different sources as well as from literature and provide rapid access to it. Analysis of such data can be further utilized to obtain important insights to understand allergic reactions. Structural features of allergens contribute significantly to their allergenicity and therefore this knowledge can be employed for developing more efficient methods for allergen/allergenicity and allergic cross-reactivity prediction. Recent advances in epitope prediction methodologies focus on antibody-specific epitope prediction approaches [150]. Application of such approaches for predicting IgE-binding epitopes will be extremely important in the development of newer and effective strategies for diagnosis and treatment of allergic diseases. Allergen immunotherapy (AIT), which is an individualized and allergens-based treatment approach, has been consid-ered as a prototype of precision medicine or personalized medicine [156]. Bioinformatics has the potential to play an important role in the development of novel approaches in AIT as well as contribute for further enrichment of the field of allergen informatics. This will surely aid in gaining better understanding of allergic diseases and positively influence upcoming research in the field.

Acknowledgements

Author details

Kiran Kadam[1], Sangeeta Sawant[1], V.K. Jayaraman[2] and Urmila Kulkarni-Kale[1*]

*Address all correspondence to: urmila@bioinfo.net.in

1 Bioinformatics Centre, Savitribai Phule Pune University, Pune, India

2 Shiv Nadar University, Gautam Buddha Nagar, Uttar Pradesh, India

References

[1]   Parkin J, Cohen B. An overview of the immune system. Lancet. 2001;357:1777–1789. DOI: 10.1016/S0140-6736(00)04904-7

[2]   Kindt T, Goldsby R, Osborne B, Kuby J. Kuby Immunology. 6th ed. New York: W.H. Freeman; 2007. 574 p. ISBN: 9780716785903

[3]   Janeway CA Jr, Travers P, Walport M, Shlomchik MJ. Immunobiology: The Immune System in Health and Disease. 5th ed. New York: Garland Science; 2001. 732 p. ISBN: 9780815336426

[4]   Aalberse RC. Structural biology of allergens. J Allergy Clin Immunol. 2000;106:228–238. DOI: 10.1067/mai.2000.108434

[5]   Pawankar R. Allergic diseases and asthma: a global public health concern and a call to action. World Allergy Organ J. 2014;7:12. DOI: 10.1186/1939-4551-7-12

[6]   Bochner BS, Rothenberg ME, Boyce JA, Finkelman F. Advances in mechanisms of allergy and clinical immunology in 2012. J Allergy Clin Immunol. 2013;131:661–667. DOI: 10.1016/j.jaci.2012.12.676

[7]   Ghosh D, Gupta-Bhattacharya S. Allergen Bioinformatics: Recent Trends and Developments. In: Xia X, editor. Selected Works in Bioinformatics. Croatia: InTech; 2011. ISBN: 978-953-307-281-4. DOI: 10.5772/20269

[8]   Scheurer S, Toda M, Vieths S. What makes an allergen? Clin Exp Allergy. 2015;45:1150–1161. DOI: 10.1111/cea.12571

[9]   King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA, Thomas W. Allergen nomenclature. Int Arch Allergy Immunol. 1994;105:224–233. DOI: 10.1159/000236761

[10] King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA. Thomas W. Allergen nomenclature. Allergy. 1995;50:765–774. DOI: 10.1111/j.1398-9995.1995.tb01222.x

[11] Breiteneder H, Chapman MD. Allergen Nomenclature. In: Lockey RF, Ledford DK, editors. Allergens and Allergen Immunotherapy: Subcutaneous, sublingual and oral. 5th ed. New York: CRC Press; 2014. pp. 37–49.

[12] Lagares A, Puerta L, Caraballo L. Polymorphism in allergens. Biomedica. 2002;22:51–62.

[13] Lowenstein H. Quantitative immunoelectrophoretic methods as a tool for the analysis and isolation of allergens. Prog Allergy. 1978;25:1–62. DOI: 10.1159/000401700

[14]   Pomés A. Relevant B cell epitopes in allergic disease. Int Arch Allergy Immunol. 2010;152:1–11. DOI: 10.1159/000260078

[15]   Bufe A. Significance of IgE-binding epitopes in allergic disease. J Allergy Clin Immunol. 2001;107:219–221. DOI: 10.1067/mai.2001.112850

[16]   Aalberse RC, Crameri R. IgE-binding epitopes: a reappraisal. Allergy. 2011;66:1261–1274. DOI: 10.1111/j.1398-9995.2011.02656.x

[17]   Bannon GA, Ogawa T. Evaluation of available IgE-binding epitope data and its utility in bioinformatics. Mol Nutr Food Res. 2006;50:638–644. DOI: 10.1002/mnfr.200500276

[18]   Matsuo H, Yokooji T, Taogoshi T. Common food allergens and their IgE-binding epitopes. Allergol Int. 2015;64:332–343. DOI: 10.1016/j.alit.2015.06.009

[19]   Meno KH. Allergen structures and epitopes. Allergy. 2011;66:19–21. DOI: 10.1111/j.1398-9995.2011.02625.x

[20]   Ghate AD, Bhagwat BU, Bhosle SG, Gadepalli SM, Kulkarni-Kale UD. Characterization of antibody-binding sites on proteins: development of a knowledgebase and its applications in improving epitope prediction. Protein Pept Lett. 2007;14:531–535. DOI: 10.2174/092986607780989921

[21]   Kulkarni-Kale U, Raskar-Renuse S, Natekar-Kalantre G, Saxena SA. Antigen–antibody interaction database (AgAbDb): a compendium of antigen–antibody interactions. In: De RK, Tomar N, editors. Immunoinformatics. New York: Springer; 2014. pp. 149–164. DOI: 10.1007/978-1-4939-1115-8_8

[22]   Cavkaytar O, Akdis CA, Akdis M. Modulation of immune responses by immunotherapy in allergic diseases. Curr Opin Pharmacol. 2014;17:30–37. DOI: 10.1016/j.coph.2014.07.003

[23]   Larché M. T cell epitope-based allergy vaccines. In: Valenta R, Coffman R, editors. Vaccines against Allergies. Berlin, Heidelberg: Springer; 2011. pp. 107–119. DOI: 10.1007/82_2011_131

[24]   Prickett SR, Rolland JM, O'Hehir RE. Immunoregulatory T cell epitope peptides: the new frontier in allergy therapy. Clin Exp Allergy. 2015;45:1015–1026. DOI: 10.1111/cea.12554

[25]   Frazier A, Schulten V, Hinz D, Oseroff C, Sidney J, Peters B, Sette A. Allergy-associated T cell epitope repertoires are surprisingly diverse and include non-IgE reactive antigens. World Allergy Organ J. 2014;7:26. DOI: 10.1186/1939-4551-7-26

[26]   Vaughan K, Greenbaum J, Kim Y, Vita R, Chung J, Peters B, Broide D, Goodman R, Grey H, Sette A. Towards defining molecular determinants recognized by adaptive immunity in allergic disease: an inventory of the available data. J Allergy (Cairo). 2011;2010:628026. DOI: 10.1155/2010/628026

[27] Aalberse RC. Assessment of allergen cross-reactivity. Clin Mol Allergy. 2007;5:1. DOI: 10.1186/1476-7961-5-2

[28] Aalberse RC, Akkerdaas J, Van Ree R. Cross-reactivity of IgE antibodies to allergens. Allergy. 2001;56:478–490. DOI: 10.1034/j.1398-9995.2001.056006478.x

[29] Ferreira F, Hawranek T, Gruber P, Wopfner N, Mari A. Allergic cross-reactivity: from gene to the clinic. Allergy. 2004;59:243–267. DOI: 10.1046/j.1398-9995.2003.00407.x

[30] Brusic V, Millot M, Petrovsky N, Gendel SM, Gigonzac O, Stelman SJ. Allergen databases. Allergy. 2003;58:1093–1100. DOI: 10.1034/j.1398-9995.2003.00248.x

[31] Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucl Acids Res. 2015;43:D30–D35. DOI: 10.1093/nar/gku1216

[32] O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucl Acids Res. 2016;4:D733–D745. DOI: 10.1093/nar/gkv1189

[33] UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucl Acids Res. 2012;40:D71–D75. DOI: 10.1093/nar/gkr981

[34] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucl Acids Res. 2000;28:235–242. DOI: 10.1093/nar/28.1.235

[35] WHO/IUIS Allergen database. Available from: http://www.allergen.org [Accessed: 2016-03-07].

[36] Chapman MD, Pomés A, Breiteneder H, Ferreira F. Nomenclature and structural biology of allergens. J Allergy Clin Immunol. 2007;119:414–420. DOI: 10.1016/j.jaci.2006.11.001

[37] Radauer C, Nandy A, Ferreira F, Goodman RE, Larsen JN, Lidholm J, Pomés A, Raulf-Heimsoth M, Rozynek P, Thomas WR, Breiteneder H. Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. Allergy. 2014;69:413–419. DOI: 10.1111/all.12348

[38] Allergome. Available from: http://www.allergome.org [Accessed: 2016-03-07].

[39] Mari A, Scala E, Palazzo P, Ridolfi S, Zennaro D, Carabella G. Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. Cell Immunol. 2006;244:97–100. DOI: 10.1016/j.cellimm.2007.02.012

[40] SDAP – Structural Database of Allergenic Proteins. Available from: https://fermi.utmb.edu [Accessed: 2016-03-07].

[41] Ivanciuc O, Schein C, Braun W. SDAP: database and computational tools for allergenic proteins. Nucl Acids Res. 2003;31:359–362. DOI: 10.1093/nar/gkg010

[42] NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucl Acids Res. 2015;43:D6–D17. DOI: 10.1093/nar/gku1130

[43] Allergen Database for Food Safety (ADFS). Available from: http://allergen.nihs.go.jp/ADFS/ [Accessed: 2016-06-01].

[44] Nakamura R, Teshima R, Takagi K, Sawada J. Development of Allergen Database for Food Safety (ADFS): an integrated database to search allergens and predict allergenicity. Kokuritsu Iyakuhin Shokuhin Eisei Kenkyujo hokoku. 2004;123:32–36.

[45] AllergenOnline. Available from: http://www.allergenonline.org [Accessed: 2016-03-07].

[46] Goodman RE, Ebisawa M, Ferreira F, Sampson HA, van Ree R, Vieths S, et al. AllergenOnline: a peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. Mol Nutr Food Res. 2016;60:1183–98. DOI: 10.1002/mnfr.201500769

[47] AllFam. Available from: http://www.meduniwien.ac.at/allfam [Accessed: 2016-03-07].

[48] Radauer C, Bublin M, Wagner S, Breiteneder H. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. J Allergy Clin Immunol. 2008;121:847–852. DOI: 10.1016/j.jaci.2008.01.025

[49] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. Pfam: the protein families database. Nucl Acids Res. 2013;42:D222–D230. DOI: 10.1093/nar/gkt1223

[50] Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol. 1995;247:536–540. DOI: 10.1016/S0022-2836(05)80134-2

[51] Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucl Acids Res. 2004;32:D258–D261. DOI: 10.1093/nar/gkh036

[52] AllergenPro. Available from: http://nabic.rda.go.kr/allergen [Accessed: 2016-06-01].

[53] Kim CK, Seol YJ, Lee DJ, Jeong IS, Yoon UH, Lee JY, Lee GS, Park DS. AllergenPro: an integrated database for allergenicity analysis and prediction. Bioinformation. 2014;10:378. DOI: 10.6026/97320630010378

[54] Saha S, Bhasin M, Raghava GP. Bcipep: a database of B-cell epitopes. BMC Genomics. 2005;6:79. DOI: 10.1186/1471-2164-6-79

[55] Huang J, Honda W. CED: a conformational epitope database. BMC Immunol. 2006;7:7. DOI: 10.1186/1471-2172-7-7

[56] Sharma OP, Das AA, Krishna R, Suresh Kumar M, Mathur PP. Structural epitope database (SEDB): a web-based database for the epitope, and its intermolecular inter-

action along with the tertiary structure information. J Proteomics Bioinform. 2012;5:84–89. DOI: 10.4172/jpb.1000217

[57] The Immune Epitope Database (IEDB). 2010. Available from: http://www.iedb.org [Accessed: 2016-03-07].

[58] Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B. The immune epitope database 2.0. Nucl Acids Res. 2010;38:D854–D862. DOI: 10.1093/nar/gkp1004

[59] Vaughan K, Peters B, Larche M, Pomes A, Broide D, Sette A. Strategies to query and display allergy-derived epitope data from the immune epitope database. Int Arch Allergy Immunol. 2013;160:334–345. DOI: 10.1159/000343880

[60] Davies V, Vaughan K, Damle R, Peters B, Sette A. Classification of the universe of immune epitope literature: representation and knowledge gaps. PLoS One. 2009;4:e6948. DOI: 10.1371/journal.pone.0006948

[61] AllerBase. Available from: http://bioinfo.net.in/AllerBase/Home.html [Accessed: 2016-03-07]

[62] Brusic V, Petrovsky N, Gendel SM, Millot M, Gigonzac O, Stelman SJ. Computational tools for the study of allergens. Allergy. 2003;58:1083–1092. DOI: 10.1034/j.1398-9995.2003.00224.x

[63] Metcalfe DD, Astwood JD, Townsend R, Sampson HA, Taylor SL, Fuchs RL. Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Crit Rev Food Sci Nutr. 1996;36:165–186. DOI: 10.1080/10408399609527763

[64] FAO/WHO. Allergenicity of Genetically Modified Foods. Rome, Italy. 2001. Available from: http://www.who.int/foodsafety/publications/biotech/en/ec_jan2001.pdf [Accessed: 2016-03-07].

[65] FAO/WHO. Codex Principles and Guidelines on Foods Derived from Biotechnology. Rome, Italy: Joint FAO/WHO Food Standards Programme. 2003. Available from: ftp://ftp.fao.org/codex/Publications/Booklets/Biotech/Biotech_2009e.pdf [Accessed: 2016-03-07].

[66] Fiers MW, Kleter GA, Nijland H, Peijnenburg AA, Nap JP, Van Ham RC. Allermatch™, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. BMC Bioinformatics. 2004;5:133. DOI: 10.1186/1471-2105-5-133

[67] Zhang ZH, Koh JL, Zhang GL, Choo KH, Tammi MT, Tong JC. AllerTool: a web server for predicting allergenicity and allergic cross-reactivity in proteins. Bioinformatics. 2007;23:504–506. DOI: 10.1093/bioinformatics/btl621

[68] Verma A, Misra A, Subash S, Das M, Dwivedi P. Computational allergenicity prediction of transgenic proteins expressed in genetically modified crops. Immunopharmacol Immunotoxicol. 2011;33:410–422. DOI: 10.3109/08923973.2010.523704

[69] Hileman RE, Silvanovich A, Goodman RE, Rice EA, Holleschak G, Astwood JD, Hefle SL. Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. Int Arch Allergy Immunol. 2002;128:280–291. DOI: 10.1159/000063861

[70] Kleter GA, Peijnenburg AA. Screening of transgenic proteins expressed in transgenic food crops for the presence of short amino acid sequences identical to potential, IgE–binding linear epitopes of allergens. BMC Struct Biol. 2002;2:8. DOI: 10.1186/1472-6807-2-8

[71] Stadler MB, Stadler BM. Allergenicity prediction by protein sequence. FASEB J. 2003;17:1141–1143. DOI: 10.1096/fj.02-1052fje

[72] The web tool WebAllergen (http://weballergen.bii.a-star.edu.sg/) could not be accessed on 2016-06-01

[73] Riaz T, Hor HL, Krishnan A, Tang F, Li KB. WebAllergen: a web server for predicting allergenic proteins. Bioinformatics. 2005;21:2570–2571. DOI: 10.1093/bioinformatics/bti356

[74] Kong W, Tan TS, Tham L, Choo KW. Improved prediction of allergenicity by combination of multiple sequence motifs. In Silico Biol. 2007;7:77–86.

[75] Saha S, Raghava GP. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucl Acids Res. 2006;34:W202–W209. DOI: 10.1093/nar/gkl343

[76] Tyagi N, Farnell EJ, Fitzsimmons CM, Ryan S, Tukahebwa E, Maizels RM, Dunne DW, Thornton JM, Furnham N. Comparisons of allergenic and metazoan parasite proteins: allergy the price of immunity. PLoS Comput Biol. 2015;11:e1004546. DOI: 10.1371/journal.pcbi.1004546

[77] Kadam K, Sawant S, Kulkarni-Kale U, Jayaraman VK. Prediction of protein function based on machine learning methods: an overview. In: iConcept Press, editor. Genomics III - Methods, Techniques and Applications. Hong Kong: iConcept Press; 2014 . pp. 125–162. ISBN: 978-1-922227-096

[78] Zorzet A, Gustafsson M, Hammerling U. Prediction of food protein allergenicity: a bioinformatic learning systems approach. In Silico Biol. 2002;2:525–534.

[79] Soeria-Atmadja D, Zorzet A, Gustafsson MG, Hammerling U. Statistical evaluation of local alignment features predicting allergenicity using supervised classification algorithms. Int Arch Allergy Immunol. 2004;133:101–112. DOI: 10.1159/000076382

[80] Li KB, Issac P, Krishnan A. Predicting allergenic proteins using wavelet transform. Bioinformatics. 2004;20:2572–2578. DOI: 10.1093/bioinformatics/bth286

[81] Dimitrov I, Naneva L, Bangov I, Doytchinova I. Allergenicity prediction by artificial neural networks. J Chemom. 2014;28:282–286. DOI: 10.1002/cem.2597

[82] AllerTOP. Available from: http://www.ddg-pharmfac.net/AllerTOP [Accessed: 2016-03-07].

[83] Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v. 2—a server for in silico prediction of allergens. J Mol Model. 2014;20:1–6. DOI: 10.1007/s00894-014-2278-5

[84] AlgPred. Available from: http://www.imtech.res.in/raghava/algpred/ [Accessed: 2016-03-07].

[85] The web tool EVALLER (http://bioinformatics.bmc.uu.se/evaller.html) could not be accessed on 2016-06-01

[86] Barrio AM, Soeria-Atmadja D, Nister A, Gustafsson MG, Hammerling U, Bongcam-Rudloff E. EVALLER: a web server for in silico assessment of potential protein allergenicity. Nucl Acids Res. 2007;35:W694–W700. DOI: 10.1093/nar/gkm370

[87] The web tool AllerTool (http://research.i2r.a-star.edu.sg/AllerTool/) could not be accessed on 2016-06-01

[88] AllerHunter. Available from: http://tiger.dbs.nus.edu.sg/AllerHunter/ [Accessed: 2016-03-07].

[89] Muh HC, Tong JC, Tammi MT. AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins. PLoS One. 2009;4:e5861. DOI: 10.1371/journal.pone.0005861

[90] The web tool APPEL (http://jing.cz3.nus.edu.sg/cgi-bin/APPEL) could not be accessed on 2016-06-01

[91] Cui J, Han LY, Li H, Ung CY, Tang ZQ, Zheng CJ, Cao ZW, Chen YZ. Computer prediction of allergen proteins from sequence-derived protein structural and physico-chemical properties. Mol Immunol. 2007;44:514–520. DOI: 10.1016/j.molimm.2006.02.010

[92] SORTALLER. Available from: http://sortaller.gzhmu.edu.cn/ [Accessed: 2016-03-07].

[93] Zhang L, Huang Y, Zou Z, He Y, Chen X, Tao A. SORTALLER: predicting allergens using substantially optimized algorithm on allergen family featured peptides. Bioinformatics. 2012;28:2178–2179. DOI: 10.1093/bioinformatics/bts326

[94] Mohabatkar H, Mohammad Beigi M, Abdolahi K, Mohsenzadeh S. Prediction of allergenic proteins by means of the concept of Chou's pseudo amino acid composition and a machine learning approach. Med Chem. 2013;9:133–137. DOI: 10.2174/1573406411309010133

[95] PREAL. Available from: http://gmobl.sjtu.edu.cn/PREAL/index.php [Accessed: 2016-03-07].

[96] Wang J, Zhang D, Li J. PREAL: prediction of allergenic protein by maximum Relevance Minimum Redundancy (mRMR) feature selection. BMC Syst Biol. 2013;7:S9. DOI: 10.1186/1752-0509-7-S5-S9

[97] Chrysostomou C, Seker H. Prediction of protein allergenicity based on signal-processing bioinformatics approach. In: Proceedings of the Engineering in Medicine and Biology Society (EMBC), 36th Annual International Conference of the IEEE 2014. Chicago: IEEE; 2014. pp. 808–811. DOI: 10.1109/EMBC.2014.6943714

[98] Allerdictor. Available from: http://allerdictor.vbi.vt.edu/ [Accessed: 2016-03-07].

[99] Dang HX, Lawrence CB. Allerdictor: fast allergen prediction using text classification techniques. Bioinformatics. 2014;30:1120–1128. DOI: 10.1093/bioinformatics/btu004

[100] Wang J, Yu Y, Zhao Y, Zhang D, Li J. Evaluation and integration of existing methods for computational prediction of allergens. BMC Bioinformatics. 2013;14:S1. DOI: 10.1186/1471-2105-14-S4-S1

[101] proAP. Available from: http://gmobl.sjtu.edu.cn/proAP/main.html [Accessed: 2016-03-07].

[102] AllergenFP. Available from: http://ddg-pharmfac.net/AllergenFP/ [Accessed: 2016-03-07].

[103] Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics. 2014;30:846–851. DOI: 10.1093/bioinformatics/btt619

[104] Bragin AO, Demenkov PS, Kolchanov NA, Ivanisenko VA. Accuracy of protein allergenicity prediction can be improved by taking into account data on allergenic protein discontinuous peptides. J Biomol Struct Dyn. 2013;31:59–64. DOI: 10.1080/07391102.2012.691362

[105] Vijayakumar S, Lakshmi PTV. A fuzzy inference system for predicting allergenicity and allergic cross-reactivity in proteins. In: Proceedings of the Bioinformatics and Biomedicine (BIBM), IEEE International Conference; 18–21 December 2013. Shanghai: IEEE; 2013. pp. 49–52. DOI: 10.1109/BIBM.2013.6732458

[106] FuzzyApp. Available from: http://fuzzyapp.bicpu.edu.in/fuzzyapp.php [Accessed: 2016-03-07].

[107] Saravanan V, Lakshmi PT. Fuzzy logic for personalized healthcare and diagnostics: FuzzyApp—a fuzzy logic based allergen-protein predictor. Omics. 2014;18:570–581. DOI: 10.1089/omi.2014.0021

[108] Jiang B, Qu H, Hu Y, Ni T, Lin Z. Computational analysis of the relationship between allergenicity and digestibility of allergenic proteins in simulated gastric fluid. BMC Bioinformatics. 2007;8:375. DOI: 10.1186/1471-2105-8-375

[109] Yang X, Yu X. An introduction to epitope prediction methods and software. Rev Med Virol. 2009;19:77–96

[110] Wang HW, Pai TW. Machine learning-based methods for prediction of linear B-cell epitopes. Methods Mol Biol. 2014;1184:217–236. DOI: 10.1007/978-1-4939-1115-8_12

[111] Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins. 2006;65:40–48. DOI: 10.1002/prot.21078

[112] Shen W, Cao Y, Cha L, Zhang X, Ying X, Zhang W, Ge K, Li W, Zhong L. Predicting linear B-cell epitopes using amino acid anchoring pair composition. BioData Min. 2015;8:14. DOI: 10.1186/s13040-015-0047-3

[113] EL-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. J Mol Recognit. 2008;21:243–255. DOI: 10.1002/jmr.893

[114] Saha S, Raghava GP. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In: Nicosia G, Cutello V, Bentley P, Timmis J, eds. Artificial immune systems. Berlin: Springer; 2004 p. 197–204. DOI: 10.1007/978-3-540-30220-9_16

[115] Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006;2:1. DOI: 10.1186/1745-7580-2-2

[116] Gao J, Faraggi E, Zhou Y, Ruan J, Kurgan L. BEST: improved prediction of B-cell epitopes from antigen sequences. PLoS One. 2012;7:e40104. DOI: 10.1371/journal.pone.0040104

[117] Wee JL, Simarmata D, Kam YW, Ng FL, Tong JC. SVM-based prediction of linear B-cell epitopes using Bayes Feature Extraction. BMC Genomics. 2010;11:1. DOI: 10.1186/1471-2164-11-S4-S21

[118] Sweredoski MJ, Baldi P. COBEpro: a novel system for predicting continuous B-cell epitopes. Protein Eng Des Sel. 2009;22:113–120. DOI: 10.1093/protein/gzn075

[119] Lian Y, Ge M, Pan XM. EPMLR: sequence-based linear B-cell epitope prediction method using multiple linear regression. BMC Bioinformatics. 2014;15:414. DOI: 10.1186/s12859-014-0414-y

[120] Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, Lundegaard C, Sette A, Lund O, Bourne PE, Nielsen M. Immune epitope database analysis resource. Nucl Acids Res. 2012;40:W525–W530. DOI: 10.1093/nar/gks438

[121] Singh H, Ansari HR, Raghava GP. Improved method for linear B-cell epitope prediction using antigen's primary sequence. PLoS One. 2013;8:e62216. DOI: 10.1371/journal.pone.0062216

[122] Yao B, Zhang L, Liang S, Zhang C. SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. PLoS One. 2012;7:e45152. DOI: 10.1371/journal.pone.0045152

[123]   Kolaskar AS, Kulkarni-Kale U. Prediction of three-dimensional structure and mapping of conformational epitopes of envelope glycoprotein of Japanese encephalitis virus. Virology. 1999;261:31–42. DOI: 10.1006/viro.1999.9859

[124]   Kulkarni-Kale U, Bhosle S, Kolaskar AS. CEP: a conformational epitope prediction server. Nucl Acids Res. 2005;33:W168–W171. DOI: 10.1093/nar/gki460

[125]   Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput Biol. 2012;8:e1002829. DOI: 10.1371/journal.pcbi.1002829

[126]   Qi T, Qiu T, Zhang Q, Tang K, Fan Y, Qiu J, Wu D, Zhang W, Chen Y, Gao J, Zhu R. SEPPA 2.0—more refined server to predict spatial epitope considering species of immune host and subcellular localization of protein antigen. Nucl Acids Res. 2014;42:W59–W63. DOI: 10.1093/nar/gku395

[127]   Sweredoski MJ, Baldi P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. Bioinformatics. 2008;24:1459–1460. DOI: 10.1093/bioinformatics/btn199

[128]   Giacò L, Amicosante M, Fraziano M, Gherardini PF, Ausiello G, Helmer-Citterich M, Colizzi V, Cabibbo A. B-Pred, a structure based B-cell epitopes prediction server. Adv Appl Bioinform Chem. 2012;5:11–21. DOI: 10.2147/AABC.S30620

[129]   Ansari HR, Raghava GP. Identification of conformational B-cell epitopes in an antigen from its primary sequence. Immunome Res. 2010;6:6. DOI: 10.1186/1745-7580-6-6

[130]   Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, Peters B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. 2008;9:1. DOI: 10.1186/1471-2105-9-514

[131]   Rubinstein ND, Mayrose I, Martz E, Pupko T. Epitopia: a web-server for predicting B-cell epitopes. BMC Bioinformatics. 2009;10:287. DOI: 10.1186/1471-2105-10-287

[132]   Liang S, Zheng D, Standley DM, Yao B, Zacharias M, Zhang C. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. BMC Bioinformatics. 2010;11:381. DOI: 10.1186/1471-2105-11-381

[133]   Liang S, Zheng D, Zhang C, Zacharias M. Prediction of antigenic epitopes on protein surfaces by consensus scoring. BMC Bioinformatics. 2009;10:302. DOI: 10.1186/1471-2105-10-302

[134]   Zheng W, Ruan J, Hu G, Wang K, Hanlon M, Gao J. Analysis of conformational B-cell epitopes in the antibody–antigen complex using the depth function and the convex Hull. PLoS One. 2015;10:e0134835. DOI: 10.1371/journal.pone.0134835

[135]   AgAbDb. Available from: http://bioinfo.net.in/AgAbDb.htm [Accessed: 2016-03-07].

[136] Weber CA, Mehta PJ, Ardito M, Moise L, Martin B, De Groot AS. T cell epitope: friend or foe? Immunogenicity of biologics in context. Adv Drug Deliv Rev. 2009;61:965–976. DOI: 10.1016/j.addr.2009.07.001

[137] Tong JC, Tan TW, Ranganathan S. Methods and protocols for prediction of immunogenic epitopes. Brief Bioinform. 2007;8:96–108. DOI: 10.1093/bib/bbl038

[138] Ponomarenko JV, Van Regenmortel MH. B cell epitope prediction. In: Gu J, Bourne PE, editors. Structural bioinformatics. New York: John Wiley & Sons; 2009. pp. 849–879.

[139] Desai DV, Kulkarni-Kale U. T-cell epitope prediction methods: an overview. In: De RK, Tomar N, editors. Immunoinformatics. New York: Springer; 2014. pp. 333–364. DOI: 10.1007/978-1-4939-1115-8_19

[140] Bhasin M, Raghava GP. Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine. 2004;22:3195–3204. DOI: 10.1016/j.vaccine.2004.02.005

[141] Doytchinova IA, Guan P, Flower DR. EpiJen: a server for multistep T cell epitope prediction. BMC Bioinformatics. 2006;7:131. DOI: 10.1186/1471-2105-7-131

[142] Dimitrov I, Garnev P, Flower DR, Doytchinova I. EpiTOP—a proteochemometric tool for MHC class II binding prediction. Bioinformatics. 2010;26:2066–2068. DOI: 10.1093/bioinformatics/btq324

[143] Stranzl T, Larsen MV, Lundegaard C, Nielsen M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics. 2010;62:357–368. DOI: 10.1007/s00251-010-0441-4

[144] Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics. 2013;65:711–724. DOI: 10.1007/s00251-013-0720-y

[145] Oyarzún P, Ellis JJ, Bodén M, Kobe B. PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity. BMC Bioinformatics. 2013;14:52. DOI: 10.1186/1471-2105-14-52

[146] Schuler MM, Nastke MD, Stevanović S. SYFPEITHI: database for searching and T-cell epitope prediction. Methods Mol Biol. 2007;409:75–93. DOI: 10.1007/978-1-60327-118-9_5

[147] Zhang L, Chen Y, Wong HS, Zhou S, Mamitsuka H, Zhu S. TEPITOPEpan: extending TEPITOPE for peptide binding prediction covering over 700 HLA-DR molecules. PLoS One. 2012;7:e30483. DOI: 10.1371/journal.pone.0030483

[148] Donnes P, Kohlbacher O. Integrated modeling of the major events in the MHC class I antigen processing pathway. Protein Sci. 2005;14:2132–2140. DOI: 10.1110/ps.051352405

[149] Farrell D, Gordon SV. Epitopemap: a web application for integrated whole proteome epitope prediction. BMC Bioinformatics. 2015;16:221. DOI: 10.1186/s12859-015-0659-0

[150] Sela-Culang I, Ofran Y, Peters B. Antibody specific epitope prediction—emergence of a new paradigm. Curr Opin Virol. 2015;11:98–102. DOI: 10.1016/j.coviro.2015.03.012

[151] Tong JC, Tammi MT. Prediction of protein allergenicity using local description of amino acid sequence. Front Biosci. 2007;13:6072–6078. DOI: 10.2741/3138

[152] Dall'Antonia F, Gieras A, Devanaboyina SC, Valenta R, Keller W. Prediction of IgE-binding epitopes by means of allergen surface comparison and correlation to cross-reactivity. J Allergy Clin Immunol. 2011;128:872–879. DOI: 10.1016/j.jaci.2011.07.007

[153] Lollier V, Denery-Papini S, Brossard C, Tessier D. Meta-analysis of IgE-binding allergen epitopes. Clin Immunol. 2014;153:31–39. DOI: 10.1016/j.clim.2014.03.010

[154] LocAllEpi. Available from: http://wwwappli.nantes.inra.fr:8180/LocAllEpi/ [Accessed: 2016-03-07].

[155] Aalberse RC, Stadler BM. In silico predictability of allergenicity: from amino acid sequence via 3-D structure to allergenicity. Mol Nutr Food Res. 2006;50:625–627. DOI: 10.1002/mnfr.200500270

[156] Canonica GW, Bachert C, Hellings P, Ryan D, Valovirta E, Wickman M, De Beaumont O, Bousquet J. Allergen immunotherapy (AIT): a prototype of precision medicine. World Allergy Organ J. 2015;8:31. DOI: 10.1186/s40413-015-0079-7