Review article
Section: Food Technology

Databases and Associated Bioinformatic Tools in Studies of Food Allergens, Epitopes and Haptens – a Review

Justyna Bucholska*, Piotr Minkiewicz, Małgorzata Darewicz, Anna Iwaniak

Department of Food Biochemistry, Faculty of Food Science,
University of Warmia and Mazury in Olsztyn, Pl. Cieszyński 1, 10–726 Olsztyn, Poland

Key words: allergy, bioinformatics, epitope, hapten, database

Allergies and/or food intolerances are a growing problem of the modern world. Difficulties associated with the correct diagnosis of food allergies result in the need to classify the factors causing allergies and allergens themselves. Therefore, internet databases and other bioinformatic tools play a special role in deepening knowledge of biologically-important compounds. Internet repositories, as a source of information on different chemical compounds, including those related to allergy and intolerance, are increasingly being used by scientists. Bioinformatic methods play a significant role in biological and medical sciences, and their importance in food science is increasing. This study aimed at presenting selected databases and tools of bioinformatic analysis useful in research on food allergies, allergens (11 databases), epitopes (7 databases), and haptens (2 databases). It also presents examples of the application of computer methods in studies related to allergies.

INTRODUCTION

Allergens are listed as one of the most important risk factors related to the impact of food on the body [Ameratunga *et al.*, 2016; Borchers *et al.*, 2010; Di Costanzo *et al.*, 2016; Jędrychowski *et al.*, 2008; Jiménez-Saiz *et al.*, 2015]. Food hypersensitivity associated or non-associated with immune mechanisms is called allergy and food intolerance, respectively. The increasing incidence of allergy and/or intolerance is the problem of the modern civilization. Development of allergy and food intolerance is stimulated by environmental factors, such as environmental contamination, as well as predispositions of the organism [Sicherer, 2011]. Factors resulting from the progress of civilization have an impact on the quality of the food consumed. This phenomenon, in turn, results in the emergence of new causes of food allergies and, in consequence, an increase in the number of new allergens. Necessity to classify the causative agents of allergies and allergens themselves is extremely important. This is related to problems associated with the correct diagnosis of food allergies, and very often non-specific symptoms of this disease. Information technologies can be applied to extend knowledge about biologically-important compounds. Among them, a special role is played by internet databases, which are repositories of information on, *inter alia*, compound structure, function, nomenclature, taxonomy, and physicochemical properties [Scalbert *et al.*, 2011]. For this reason, online da-tabases and bioinformatic analysis tools (*in silico*) cooperating with them are increasingly becoming new research tools and are extensively addressed in many studies [de la Iglesia *et al.*, 2013; Hayes *et al.*, 2015; Kadam *et al.*, 2016; Minkiewicz *et al.*, 2009; 2013; Sircar *et al.*, 2014]. This trend is also observed in the case of databases related to food allergies and associated with them allergens of different origin. A website approved by the World Health Organization (WHO) and Commission for Allergen Nomenclature of the International Union of Immunological Societies (IUIS), as well as Allergome and AllFam databases are examples of bioinformatic tools that provide information about allergies and food allergens. Bioinformatic analyses (computer-aided) associated with allergenicity of food constituents require the access to reliable and current sources of information on the sequence and structure of known allergens [Brusic *et al.*, 2003]. An important issue related to research on allergy is, *i.a.*, the determination of fragments recognized in the structure of the allergenic protein and responsible for inducing immuno-related food hypersensitivity reactions. Such fragments are referred to as epitopes [Arnon & Van Regenmortel, 1992; Wróblewska *et al.*, 2007].

In addition to information about allergies, allergens and their epitopes, there are internet repositories containing data related to haptens. These are compounds that do not induce an immune response themselves until they form a complex with an immunogenic carrier, for example a protein. Therefore, haptens may play an important role in the immune response against a number of compounds [Goodrow *et al.*, 1990; Singh *et al*., 2004].

* Corresponding Author: Tel.: +48 89 523 35 85; Fax: +48 523 39 45;
E-mail: justyna.bucholska@uwm.edu.pl (Dr. J. Bucholska)

The aim of the study was to present selected databases and bioinformatic analysis tools useful in research on food allergies, allergens, epitopes, and haptens. The study shows also examples of the application of computer methods in research related to allergies.

DATABASES OF ALLERGENS

Allergome [Mari *et al.*, 2009] is a database of allergens, created based on the literature published since the early sixties. These data are continually completed and updated. The Allergome database provides information on allergens that induce the symptoms of many diseases, such as asthma, atopic dermatitis, anaphylaxis, allergy, rhinitis, urticaria, conjunctivitis, as well as data on the sources of allergens. Allergen information shared in the Allergome database is derived from international scientific journals and internet resources. The list of allergens stored in the Allergome database includes such information about them as the source of origin (plants or animals), routes of exposure (*e.g.*, digestive tract, direct contact, inhalation), or source of occurrence (pollen, spores, seeds, fruits, venoms, hair, *etc.*). The Allergome database does not contain compounds that do not cause immune-related diseases (celiac disease) nor immunoglobulin-E dependent allergy (*e.g.*, drug allergy). This database does not contain information on the substances causing intolerance, including food, or inducing pseudo-allergic reactions. Most allergens present in the Allergome database have a name given in accordance with the nomenclature of the International Union of Immunological Societies (IUIS). The Allergome database is a repository of all allergens available on the IUIS website, as well as those which this organization does not include. Allergome also comprises information on epitopes or peptides of allergenic particles and experimentally determined allergens. Allergenic particles are divided into 273 groups with respect to their biological functions. The Allergome database has several modules, *e.g.*: RefArray (The Allergome Reference Archive containing all the literature papers processed excluding PubMed citations), ReTiME (The Real Time Monitoring of IgE sensitization to acquire and store real time data), InterAll (the allergy electronic record available for clinical centers and research laboratories), and AllegomeCustomer (useful informative module to patients with food allergy problems). Since 2007, the Allergome database has been interconnected to AllFam database [Radauer *et al.*, 2008], which includes allergen classifications for protein families. The access to and use of the database is free [Mari *et al.*, 2006; 2009].

Records of allergenic proteins in the UniProt database contain links to records of the same proteins in the Allergome. There is also search engine enabling screening of the database using sequences of short peptides (*e.g.* identified *via* sequencing using mass spectrometry). Allergome is very useful in proteomic analyses due to the tools provided, *e.g.* search option named Allergome Aligner. Proteomic methods, such as mass spectrometry together with sequence alignments, are used in the investigations on food proteins, including allergens [Ahsan *et al.*, 2016; Koeberl *et al.*, 2014; Minkiewicz *et al.*, 2015; Ortea *et al.*, 2016].

The AllAllergy website is one of the most comprehensive resources of information on allergens and food intolerances available in the internet [Brusic, 2006; Gendel, 2009]. It contains a database of allergens and other substances, such as additives and/or preservatives, responsible for allergic or intolerance reactions. Searching the AllAllergy database for compounds of interest is based on the entry of the name of a specific allergen or a category to which it is assigned. Classification categories of individual allergens include: additives and preservatives; cereals; drugs; eggs and poultry; fish and seafood (*e.g.*, crustaceans); other foods; fruits; grass pollen; insects and their venom; meat, eggs, milk and cheese; mites; molds and bacteria; nuts, seeds and grains; molluscs and crustaceans; spices and herbs; tree pollen; vegetables; and weed pollen [Brusic, 2006; Gendel, 2009].

The Allergen Online database [Goodman *et al.*, 2016] is a separate part of the FARRP website, *i.e.*, the Food Allergy Research and Resource Program. It is systematically updated. The Allergen Online database contains information on allergenic proteins and their potential cross-reactions. The purpose of this database is to evaluate the safety of proteins that can be incorporated into food by genetic engineering or by methods used in food processing. An additional function of this database is the identification of proteins that may be subject to additional tests, such as immunoglobulin E binding (IgE) in serum or histamine release. According to recent data (May 2017), Allergen Online contained 2,035 protein allergen sequences divided into 808 taxa, which comprised experimentally-confirmed allergenicity and potential allergens. The allergen search process in this database is simple, as it only requires an entry of the analyzed sequence in the form of a query. It, however, needs to be entered in the FASTA format [Pearson, 1988]. An additional database located on the FARRP website is a separate bioinformatic tool allowing to evaluate whether a protein belonging to the same subfamily as wheat gliadins, is similar enough to gluten to be able to induce symptoms typical of celiac disease [Goodman *et al.*, 2016; Siruguri *et al.*, 2015].

The Allergen Database for Food Safety (ADFS) [Nakamura *et al.*, 2005] is a database providing information on allergenic proteins and allergenic low molecular weight compounds. This database also contains links to other sites, such as allergen databases and bioinformatic tools useful for their analysis [Nakamura *et al.*, 2005].

The list of allergenic proteins collected in ADFS is based on literature data and resource information from the Allergen Online database. Allergens included in Allergome Online and ADFS databases were subject to verification by experts from the field of allergology. The data on allergens available in ADFS, such as protein sequence or higher order structure, are completed with additional information from online repositories, such as UniProt [The Uniprot Consortium, 2015]. Allergens collated in the Allergen Database for Food Safety can be searched according to the following criteria: alphabetical order, name, category, keywords, protein and epitope sequence, and predicted allergenicity [Nakamura *et al.*, 2005]. In the case of low molecular weight allergenic compounds available in ADFS, their data come from reviews of information about them in the AllAllergy database. Allergenic

low molecular weight compounds are searched on the basis of their names (in alphabetical order), keywords, and structure. Currently (May 2017), the ADFS database contains 2,028 allergen sequences, 143 allergen epitopes described in the literature, 128 allergen structures, and 129 allergens binding through sugar residues. In addition, ADFS provides information about registered users and date of the last database update. Although the use of this database is free, it requires user registration [Nakamura *et al.*, 2005].

The IUIS Allergen Nomenclature is a database dedicated to the nomenclature of allergens approved by the WHO/IUIS. It was established in 1984 by experts in the field of allergens and their structure and function, as well as specialists representing molecular biology and bioinformatics to develop a nomenclature system of allergens. Responsibility and competence of IUIS rely on maintaining and developing a unique, unambiguous and systematic nomenclature of proteins responsible for the occurrence of allergies. In addition, the IUIS Executive Committee reviews the submissions of new allergens, decides about their nomenclature and assigns the correct nomenclature to them. IUIS Allergen Nomenclature contains information on officially recognized and approved allergens. At present (May 2017), this database includes the following categories of organisms along with the number of allergens shown in brackets belonging to a given category: *Arthropods* (252), *Chordates* (82), *Cnidarians* (1), *Molluscs* (4), *Nematodes* (18), *Ascomycota* (86), *Basidiomycota* (23), *Zygomycota* (2), *monocotyledonous plants* (97), *dicotyledonous plants* (292), and *coniferous plants* (18). To find information in the database, the name (common or systematic) of the allergen or source of its origin should be used as a query [Radauer *et al.*, 2014].

The InformAll database, curated by an international team coordinated by prof. E. N. Clare Mills (University of Manchster, UK) and described in a review article by Gendel [2009], contains data on, *inter alia*, food ingredients, which may be responsible for allergy induction. Its goal is to provide users the access to the reliable source of information on food allergens of plant and animal origin. This database contains a list of allergens (*i.e.*, allergens assigned to individual organisms, of 84 listed in the database) along with their characteristics consisting of the common and systematic name, the occurrence as well as useful information for persons suffering from allergies [Gendel, 2009].

The Structural Database of Allergenic Proteins (SDAP) [Ivanciuc *et al.*, 2009] is a database of allergenic protein structures integrated with various computational tools that can be helpful in the analysis of allergen structure. This database is an important tool for bioinformatic analysis used to predict cross-reactivity between allergens and IgE binding by food proteins. The SDAP website contains an alphabetical list of allergens. Looking up an allergen in the database is possible by its scientific name or origin (scientific or common name of an organism, from which the particular allergenic protein is derived), description or type (*e.g.*, pollen of weeds, grasses and trees, dust mites, animals, fungi, insects, food and others). Allergen data search is carried out within all allergens collected in the SDAP website or in the International Union of Immunological Societies database (IUIS). In May 2017,

the SDAP database contained 1,526 allergens. In addition, SDAP contains 1,312 protein sequences, 458 three-dimensional allergen models, and 29 allergens with IgE epitopes responsible for allergy development.

The BIOPEP database was established at the Department of Food Biochemistry at the Faculty of Food Science, UWM in Olsztyn. It contains four databases integrated with each other: proteins (different origins), biologically-active peptides, sensory peptides and amino acids, and allergenic proteins with their epitopes. Information concerning the allergenic proteins is available under the tab "*Allergenic proteins and their epitopes*" [Dziuba *et al.*, 2013]. Each of the protein stored therein is described by its name, identification number, source of origin, number of amino acid residues in the protein and its structure, monoisotopic and chemical mass. Literature sources on allergenic proteins and references to other online databases complete the information on allergenic proteins. The BIOPEP database, in addition to the structure of allergenic proteins (amino acid sequence), contains also structures of known epitopes determined experimentally and by *in silico* methods (tab: "*Additional information*"). The location of the epitope in the polypeptide chain of allergenic proteins is provided as well. According to recent data (May 2017), the BIOPEP database includes a total of 135 amino acid sequences of allergenic proteins and their epitopes, if known [Dziuba *et al.*, 2013]. The database also provides annotations concerning the occurrence of linear epitopes in more than one protein. This is a commonly occurring phenomenon [Minkiewicz *et al.*, 2015]. The database also enables performing calculations, such as epitope profiles, *i.e.*, determining the location of epitopes in a protein sequence, known as *in silico* epitope mapping [Leung *et al.*, 2014]. The database may be suitable for the determination of epitope profiles for potential allergens belonging to families grouping proteins with amino acid sequences containing many common fragments such as caseins [Erhardt *et al.*, 2016], tropomyosins [Marti *et al.*, 2007] or prolamins [Minkiewicz *et al.*, 2015]. For protein families with highly variable sequences such as fish parvalbumins [Carrera *et al.*, 2006], the analysis of common epitopes *in silico* using the BIOPEP database may provide no results.

AllFam [Radauer *et al.*, 2008] is a database of allergen families. Their classification and affiliation to a particular family of proteins is based on information present in the Allergome [Mari *et al.*, 2006, 2009] and Pfam databases (http://pfam.sanger.ac.uk/) [Finn *et al.*, 2010]. AllFam contains information about proteins, which allergenicity was confirmed experimentally (*e.g.*, proteins binding human IgE), but provide no information on allergens with undefined route of exposure [Radauer *et al.*, 2008].

The AllFam database can be helpful in analyzing allergies with respect to searching allergens that cause cross reactions [Radauer *et al.*, 2008].

AllerBase is another repository of information on allergens [Kadam *et al.*, 2016]. This database provides comprehensive information on allergens from different sources with experimentally-confirmed allergenicity. It also contains much additional information, such as data on IgE epitopes, experimental allergen validation (tests), cross-reactivity. Currently

(May 2017), AllerBase contains data on 2,340 allergens. Important elements of this database are tools allowing to visualize allergen structure and the so-called index of completeness, which indicates the availability of data for an allergen, *e.g.*, concerning its experimental validation. This database also contains links to other databases devoted to allergens, *e.g.*, Allergome, AllFam or IUIS. Searching for information about the allergen is possible based on its name, type of the source organism, and presence in food. It is also possible to search for different isoforms of particular allergens, IgE epitopes, antibodies or allergen cross-reactions [Kadam *et al.*, 2016].

Information concerning the allergenic food proteins can also be found in the UniProt database [The Uniprot Consortium 2015]. The list of allergens comprises those that are related to IgE-dependent allergies [King *et al.*, 1994]. Each allergen is described according to the specific IUIS nomenclature. The source of origin and the Uniprot database number of the allergen are also assigned in addition to its symbol. In the case when the names of two allergens are identical, then extra characters are added to discriminate between them (one or more). According to the data of May 2017, the UniProt/ Swiss-Prot database dedicated to allergens contained 435 allergenic sequences.

Databases of allergens, epitopes, and haptens described in this work are shown in Table 1.

DATABASES OF EPITOPES AND HAPTENS

B-cell and T-cell lymphocytes are important in the human immune system. Recognizing antigen epitopes is the function of B-cells. In some cases, T-cells play an important role in cellular immunity through the mediation of antigenic peptides interacting with T-cells. An antigen is a protein molecule, which is not presented to immune cells as a whole, particularly to B- and T-cells. It is converted to antigenic peptides, which in turn have the ability to stimulate the host's immune system. An epitope (antigenic determinant) is part of the antigen recognized by a specific receptor of the immune system [Gołąb *et al.*, 2010]. Due to the important role of antigens in the human immune system, databases dedicated to epitopes are becoming important bioinformatic tools in the characterization of allergen fragments.

Haptens play an important role in many processes, including immune response induced by drugs, endogenous compounds (in the case of autoimmune diseases), or allergies against xenobiotics. They are compounds which do not induce an immune response themselves until they form an immunogenic complex with a carrier. Most carriers are secretory proteins or cell surface proteins, and certain polymers. A hapten and a carrier may bind with each other through covalent or non-covalent bonds [Shreder, 2000].

The Immune Epitope Database (IEDB) [Vita *et al.*, 2015] is a database containing information on the relationships between antibodies and T-cell epitopes of humans, other primates, rodents and other animals. Currently (May 2017), IEDB contains information on 296,970 peptide epitopes, 2,503 non-peptide epitopes, 318,176 T-cells, 393,026 B-cells, and 597,727 ligands of MHC molecules (Major Histocompatibility Complex). The number of individual epitopes

and other data available in IEDB are constantly updated. Searching for information is based on parameters entered by the user that include: epitope structure, source of origin, type of immune disease, and immune response recognition [Vita *et al.*, 2015]

The Structural Epitope Database (SEDB) collects data on the three-dimensional structure of epitopes and information about the intermolecular interactions between the antigen and antibody. SEDB contains epitopes defined as: linear epitopes (419), non-linear epitopes (126), B-cell epitopes (272), T-cell epitopes (49), and epitopes binding MHC molecules (299). Their current number is provided in parentheses (May 2017). Epitopes may be searched in the SEDB based on the following information: keywords, ID number in the PDB database, sequence, author's name or study title on this subject (reference publications). The latest version of the SEDB database also contains immunogenic epitopes, the three-dimensional structure of which is stored in the PDB database [Sharma *et al.*, 2012].

The B-CELL Epitope Interaction Database (BEID) is a database of interactions of B-cells and epitopes [Tong *et al.*, 2008]. It comprises structures of antigen-immunoglobulin (Ig) type complexes and their characteristics. BEID is supplemented with data from the Protein Data Bank (PDB), which are then verified in terms of formation of the appropriate antigen-immunoglobulin complex interactions. Currently (May 2017), the BEID database contains structures of 164 antigens, 126 immunoglobulins, and 189 antigen-immunoglobulin interaction complexes. Searching the data in the BEID database can be performed based on the information regarding the antigen-antibody complex and such parameters as: the origin of a particular antibody, its name, antibody isotope (*e.g.* IgE), ligand name and category. It is also possible to search for information based on the data from the PDB (*i.e.*, number, type of ligand), as well as the simultaneous search for information based on several of the above categories [Tong *et al.*, 2008].

Bcipep [Saha *et al.*, 2005] is a library of peptides playing a role in humoral immunity of the organism. The database contains 3,031 peptide sequences (May 2017). Peptides stored in the Bcipep database are characterized by different immunogenicity. Information about peptides in the Bcipep database helps to improve methods for prediction of B-cell epitopes, production of synthetic vaccines, and diagnostics of many diseases. Peptides described in this database induce the production of antibodies that bind to the antigen and are responsible for body defense. The database provides the user with information regarding the following: immunogenic peptide sequences given in the form of a single letter code, source protein from which the antigenic peptide is derived, antigenic peptide position in a protein sequence, and a pathogen group to which the source protein or an antigen belongs. Searching peptides collected in Bcipep can be done on the basis of their amino acid sequences. There is a possibility to expand the results of the query by the following categories of information about the peptides: source of origin, immunogenicity, which is a measure of immunogenic peptide activity, antigen structure, produced monoclonal or specific antibodies, type of model organism (*e.g.* mice used for the analysis of immu-

TABLE 1. Databases of allergens, epitopes, and haptens (accessed: May 2017).

Name	URL	Reference	Number of citations		
			Web of Science	Scopus	Google Scholar
Allergome	http://www.allergome.org/	Mari *et al.* [2009]	9	125	628
AllAllergy	http://allallergy.net	Brusic [2006]; Gendel [2009]	0	3	827
FARRP (Allergen Online)	http://www.allergenonline.org/	Goodman *et al.* [2016]	6	14	650
ADFS	http://allergen.nihs.go.jp/ADFS/	Nakamura *et al.* [2005]	9	53	99
IUIS Allergen Nomenclature	http://www.allergen.org/	Radauer *et al.* [2014]	35	190	2570
InformAll	http://research.bmh.manchester.ac.uk/informAll/	Gendel [2009]	2	21	5585
SDAP	http://fermi.utmb.edu/SDAP/	Ivanciuc *et al.* [2009]	21	231	1
BIOPEP[1]	http://uwm.edu.pl/biochemia/index.php/pl/biopep	Dziuba *et al.* [2013]	36	177	198
Allfam	http://www.meduniwien.ac.at/allfam/	Radauer *et al.* [2008]	7	19	176
AllerBase	http://bioinfo.net.in/AllerBase/Home.html	Kadam *et al.* [2016]	0	1	10
UniProt/Swiss-Prot	http://www.uniprot.org/docs/allergen	The Uniprot Consortium [2015]	63	19	153
IEDB	http://www.iedb.org/	Vita *et al.* [2015]	97	473	824
SEDB	http://sedb.bicpu.edu.in	Sharma *et al.* [2012]	3	39	355
BEID-B-CELL Epitope Interaction Database	http://datam.i2r.a-star.edu.sg/BEID/	Tong *et al.* [2008]	0	0	0
Bcipep	http://www.imtech.res.in/raghava/bcipep/info.html	Saha *et al.* [2005]	4	117	220
MHCBN	http://www.imtech.res.in/raghava/mhcbn/index.html	Lata *et al.* [2009]; Bhasin *et al.* [2003]	5	193	371
ProPepper	https://propepper.net/	Juhász *et al.* [2015]	1	4	23
AntiJen	http://www.ddg-pharmfac.net/ddg/services.html	Blythe *et. al.* [2002]; McSparron *et al.* [2003]	4	136	220
HaptenDB	http://www.imtech.res.in/raghava/haptendb/	Singh *et al.* [2006]	1	26	48
SuperHapten	http://bioinformatics.charite.de/superhapten/	Gűnther *et al.* [2007]	1	11	135

[1]Includes citations of all subbases of BIOPEP (allergenic proteins, bioactive peptides and sensory peptides).

nogenic properties of a peptide), and experimental methods used to determine the immunogenic properties of a peptide. All peptides collected in Bcipep are also present in Swiss-Prot and PDB databases. Currently (May 2017), Bcipep contains 10,700 antigenic protein sequences and 1,216 structures of these proteins [Saha *et al.*, 2005].

The Major Histocompatibility Complex Binding, Non-binding peptides and T-cell epitopes (MHCBN) [Bhasin *et al.*, 2003; Lata *et al.*, 2009] is a database that contains detailed information on the major histocompatibility complex (MHC), non-bound peptides, and T-cell epitopes. The main purpose of the MHCBN database is to provide information for the academic community on the design of DNA vaccines. Version 4.0 of the database includes 25,860 sequences of MHC binding and non-binding peptides, about 1,052 peptides interacting with the structure referred to as a transporter associated with antigen processing (TAP). MHCBN can also be accessed using different bioinformatic tools, such as: Swiss-Prot, NCBI, PDB, and GenBank [Bhasin *et al.*, 2003; Lata *et al.*, 2009].

Propepper [Juhász *et al.*, 2015] is a database of protein epitopes, mainly prolamins of cereals and peptides. Information related to these compounds is particularly useful for users, whose interests are focused on protein chemistry, proteomics, mass spectrometry, and the development of methods applicable in, *e.g.*, food science. Currently (May 2017), Propepper contains 2,484 sequences of proteins, 37,914 peptides, 833 epitopes, 94,924 records for protein-epitope complexes, 119,069 records on the peptide-epitope matching, and 667,402 records containing data concerning digestion in protein-peptide combination. Each protein sequence in Propepper is characterized using information, such as ID number in UniProt and GenBank databases, protein sequence and its chain length, type of protein, genotype, and genome. The tab containing information on a protein-epitope complex provides additional information on the type of antibody, name of source organism and epitope sequence, location in the amino acid sequence, type of induced disease (*e.g.*, celiac disease, allergy), while the tab concerning the digestion includes information on the type of the enzyme and the se-

quence of the released peptide. The Propepper database contains also graphs illustrating the number of proteins or peptides in the database in relation to the length of the amino acid sequence, source organism, type of the manifesting disease, and enzyme used causing peptide release (digestion process). Based on the analysis of sequence frequency graphs, it can be observed that most of the proteins found in this database were present in wheat. Moreover, the highest number of epitopes originating from wheat cause celiac disease. Additionally, low-specificity chymotrypsin turned out to be the most efficient enzyme releasing the products of proteolysis [Juhász *et al.*, 2015].

The next database concerning epitopes is AntiJen [Blythe *et al.*, 2002; McSparron *et al.*, 2003]. It was elaborated at the Faculty of Pharmacy, Medical University of Sofia, and provides information on the B and T Cell epitopes, MHC ligand, TAP as well as immunological protein-protein interactions. This database includes Peptide Library, Copy Numbers, and Diffusion Coefficient data. All data found in AntiJen is derived from the experimental study publishing sources. The AntiJen contains 24,000 entries (May 2017). Searching the database for compounds of interest is based on the quick search (keyword search) or full search, *e.g.*, search for the epitope strings containing variable amino acid positions, the specific amino acid which are requested in any epitope string.

All the above-described epitope databases contain also sequences that are fragments of allergenic proteins derived from food.

HaptenDB is one of the databases containing information on haptens and their carriers [Singh *et al.*, 2006]. HaptenDB is also a database providing information on raising antibodies against a particular group of haptens, specificity and cross-reactivity of antibodies with related haptens, and the use of antibodies in constructing cost-effective and simple kits to detect them. Haptens present in HaptenDB include: vitamins, drugs, steroids, herbicides, pesticides, insecticides, hormones, dyes, toxins, explosives *etc*. The database includes 2,021 (May 2017) records containing information on antibodies or their cross-reactions in relation to one hapten. HaptenDB contains data on 1,087 haptens. These data include information about common and chemical name, molecular weight, physical and chemical properties, biological significance, and structure of haptens. One of the tools associated with HaptenDB is a search engine for hapten structure similarities. A prerequisite to search for hapten structures on the basis of similarity is a requirement to provide their chemical structure in Molfile (.mol) or Protein Data Bank (.pdb) format [Singh *et al.*, 2006].

HaptenDB consists of the following subdatabases (the type of information related to a group of compounds assigned to the appropriate subdatabase is given in parentheses [Singh *et al.*, 2006]):

– haptens (including hapten name, physical properties, molecular weight, toxicity, application, biological activity, chemical structure);
– antibodies (including hapten name, carrier name, conjugation method, antibody name);
– carriers (carrier name, carrier characteristics, sequence, physical characteristics);

– tools facilitating the search (keywords, search using the structure, overview by the name of haptens, carriers, antibodies);
– links to other websites (links to journals, links to databases).

SuperHapten is another database describing haptens [Gŭnther *et al.*, 2007] that contains information about immunogenic compounds. The chemical compounds stored in the SuperHapten database are divided into the following groups: halogen derivatives (2032), pesticides (227), herbicides (97), fungicides (43), insecticides (66), contact allergens (195), drugs (392), and natural compounds (1397). The current number of these groups (May 2017) in SuperHapten is provided in parentheses. In total, this database contains 7,257 haptens and allows searching for compounds stored in it, according to the following criteria: keywords, structure, and similarity. It also offers additional information on antibodies, carriers, and their definitions. The website provides also statistics on the use of the database and links to other sites related to antibodies, access to public library resources and other databases [Gŭnther *et al.*, 2007].

Internet repositories, as a source of information on different chemical compounds, including those related to allergy and intolerance, are increasingly being used by scientists. It is evidenced by the growing number of their citations on bibliometric pages, such as: Web of Science (https://apps.webofknowledge.com/), Scopus (https://www.scopus.com/home.uri), and Google Scholar (https://scholar.google.pl/). This trend also applies to citations of databases and bioinformatic tools dedicated to allergies. De la Iglesia and coworkers [2013] have presented number of citations in Web of Science for particular databases. More recent publications, concerning various areas of science, *e.g.,* biomedical sciences, recommend application of all three resources [Harzing & Alkangas, 2016, Trapp, 2016]. The data on this subject are summarized in Table 1. The number of citations related to the individual databases of allergen epitopes, haptens and computer programs dedicated to the analysis of these compounds was obtained using the following queries: name of a bioinformatic tool (database or program) and the word allergen/epitope/hapten. The keywords entered are provided in all search fields. Methods of data mining according to different criteria depending on the database used are summarized in Table 2. The number of citations of particular bioinformatic tools (Table 1) varies depending on the database used, *i.e.*, Web of Science, Scopus, or Google Scholar. The number of citations depends also on the time of database existence. New databases such as ProPepper or AllerBase, described in 2015 and 2016 respectively, possess less citations than the older ones. A small number of citations of any databases in Web of Science is probably caused by the fact that this resource indicates only citations in a limited number of selected journals. Scopus provides much more citations due to the fact that it summarizes much more resources (journals, books and book chapters) than the Web of Science does. For instance Scopus, provided by the Elsevier, indicates all publications of this publisher, whereas the Web of Science – only part of them. Both above databases offer similar search options allowing users to find da-

TABLE 2. Search possibilities for allergens, epitopes, and haptens in different databases (downloaded: May 2017).

No.	Database	Number of allergens/epitopes/ haptens in the database	Finding information about the allergen/epitope/hapten on the basis of:		
			allergen/epitope/ hapten name	allergen/ epitope sequence	sequence similarity
1.	Allergome	3075	+	+	+
2.	AllAllergy	1091	+	-	-
3.	FARRP (Allergen Online)	1956	+	+	+
4.	ADFS	2028	+	+	-
5.	IUIS Allergen Nomenclature	853	+	-	-
6.	InformAll	164	+	-	-
7.	SDAP	1526	+	+	+
8.	BIOPEP	135	+	-	-
9.	AllFam	936	+	-	-
10.	AllerBase	2311	+	-	-
11.	Uniprot/ SwissProt	435	+	+	-
12.	IEDB	239,253	+	+	+
14.	SEDB	1165	+	+	-
15.	BEID-B-CELL Epitope Interaction Database	479	+	-	+
16.	Bcipep	3031	-	+	+
17.	MHCBN	26913	-	+	-
18.	ProPepper	922626	-	+	-
19.	AntiJen	24000	+	-	+
20.	HaptenDB[1]	1087	+	-	+
21.	SuperHapten[1]	7734	+	-	+

[1]Hapten search is based on its name or ID number in the database. It is also possible to use a similar structure search option, *e.g.*, on the basis of hapten chemical structure, the use of molecule editor, *i.e.*, Chime, MView [Ertl, 2010]. Proteins that are hapten carriers can be searched based on the amino acid sequence.

tabase name in the title, keywords, abstracts or titles of references. Scopus and Web of Science are designed to indicate and document scientific publications. Google Scholar indicates more journals and books than Web of Science and Scopus and some categories of resources omitted by the two latter repositories. It takes into account for instance student's diploma theses uploaded in the Internet. Such dissemination of results is practiced by many universities in the world. Google Scholar provides an opportunity to find information about databases not only in strictly scientific publications but also in texts aimed at education or popularization of knowledge concerning allergies and allergens. Thousands of citations of the InformAll database in Google Scholar suggest that this database is broadly exploited not only for scientific purposes but also for education and popularization of knowledge. On the other hand, some search results concerning this database may be false positive. Google Scholar displays also texts containing word "informal" or phrase "inform all", similar to the name "InformAll". Search engines of Scopus and Web of Science accept only exact matching of words and allow avoid such mistakes. The number of cita-

tions of BIOPEP is also overestimated in all three databases as mentioned in the footnote to Table 1.

In 4 out of 11 allergen databases (Table 2), it is possible to search for information on the allergen on the basis of its name, sequence and, sequence similarity. As regards hapten databases, it is possible to search for individual compounds by their name or structure similarity.

SELECTED EXAMPLES OF THE USE OF DATABASES AND BIOINFORMATIC TOOLS IN THE RESEARCH ON ALLERGIES AND EDUCATION

One of the primary objectives of bioinformatic analysis in the field of research related to allergy and allergens is to identify protein sequences that exhibit sufficient similarity to allergens or have the potential to cause cross-reactions. According to this view, Goodman & Wise [2006] compared the sequences of proteins of rice variety Golden Rice 2 to data stored in the Allergen Online database. This way, they attempted to answer the question whether the analyzed protein sequences may be a potential source of known and previ-

ously described allergens or cause cross-reactions in people who are potential consumers of products that contain Golden Rice 2. Based on the applied criteria of sequence similarity (E similarity coefficient $<1\times10^{-7}$ and sequence identity above 50%), no structural similarities were found between the query sequences (Golden Rice 2) and a set of proteins with known allergenic properties collected in the Allergen Online database [Goodman *et al.*, 2016; Goodman & Wise, 2006].

The data contained in the SDAP database were used to analyze the evolutionary similarity between profilins of different plant species. Profilins are proteins commonly occurring in nature with a different allergenic potential primarily associated with pollen allergy. People with hay fever may also exhibit symptoms of oral allergy syndrome (OAS) after the consumption of raw vegetable foods (*e.g.*, fruits and/or vegetables). The results in the form of a taxonomic tree obtained on the basis of the data from the SDAP database showed profilins to exhibit the evolutionary similarity to sequences of bananas, melons, pineapples, sunflowers, tomatoes, celery, almonds, pears, apples, and kiwi [García & Lizaso, 2011]. The SDAP database was used to search for fragments with a sequence similar to soy protein epitopes in the sequences of other food proteins [Saeed *et al.*, 2016].

Dziuba *et al.*, [2013] used the BIOPEP database of allergenic proteins and their epitopes and associated bioinformatic tools to identify potential biomarkers present in allergenic proteins. Sequencing fragments originating from milk proteins, egg white, wheat and peanuts were proposed as biomarkers. The study used data stored in multiple databases, including allergens such as: Allergen Online, AllFam, Allergome, BIOPEP and other. The analysis included computer simulation of hydrolysis of more than 40 food proteins of known allergenicity and subsequent search among the proposed biomarkers of resulting products of proteolysis [Dziuba *et al.*, 2013]. Genotyping of biomarkers of allergenic proteins derived from fish protein was another example of the use of bioinformatic tools available in BIOPEP. The study showed that carp and herring proteins had fragments that could serve as markers common to more than one species. The same peptides were hydrolysis products of carp proteins with known sequences and herring proteins, which sequences have not yet been identified. Such peptides may be markers characteristic of protein families, but not for individual sequences [Bucholska & Minkiewicz, 2016].

Tapal and co-workers, [2016] used the BIOPEP database of allergenic proteins and their epitopes. Queries consisted of peptide sequences identified in the protein proteolysis products of oil palm *Elaeis guineensis* var. *Tenera* using mass spectrometry. The identified peptides did not contain fragments identical to known epitopes in the BIOPEP database.

The Allergome database was the source of the data that was used to find characteristic motifs related to cross-reactions between allergens in amino acid sequences of allergenic proteins (including food proteins) [Pfiffner *et al.*, 2010]. According to the definition of Liu *et al.* [2008], a motif is a reproducible pattern present in the amino acid sequences of proteins, associated with a particular biological activity.

Garino *et al.* [2016] used bioinformatic tools to study the potential allergenicity of lipid transport proteins. Part of the proteins of this family, including those derived from the raw materials used in the production of food, have not been so far experimentally tested for allergenic properties. One of the algorithms used to assess the potential allergenicity was the analysis of sequence similarity of the studied proteins and previously known allergens belonging to the same family. Data on lipid transport proteins, formerly known as allergens, was derived from the Allergome database. The study showed that each of the 28 analyzed protein sequences of different origin can be treated as a potentially allergenic protein.

Bioinformatic tools, including databases, have recently been increasingly used in education, especially in the field of biological and medical sciences [Atwood *et al.*, 2015; Brazas & Ouelette, 2016; Minkiewicz *et al.*, 2016]. The popularity of databases and computer programs in education in the agricultural sciences, including food science, is much lower than in the fields mentioned above, an example of which may be the statistics on the interests of participants of online bioinformatic courses [Ding *et al.*, 2014]. Nevertheless, the importance of bioinformatic methods in food science is increasing [Holton *et al.*, 2013], and, therefore, their spread can be expected also in the education in this field. An example of the use of the database of allergenic proteins in the education of students is an exercise introduced in 2014 at the Biochemistry course in: "*Food Science and Human Nutrition*" and "S*cience of Commodities*" specializations at the Department of Food Science, University of Warmia and Mazury in Olsztyn, Poland.

Protein sequences used in the exercise were derived from the UniProt database (http://www.uniprot.org/) [UniProt Consortium, 2015]. A class instructor assigns to students sequences of known allergens, whose properties have been confirmed experimentally on the basis of the criteria recommended by IUIS [Breiteneder & Chapman, 2014]. The sequences are analyzed by the students using the AlgPred program [Saha & Raghava, 2006], designed to predict the allergenicity of proteins. The program is based on five algorithms predicting the allergenicity, *i.e.*: search for fragments identical with known epitopes (so-called *in silico* epitope mapping), search for motifs characteristic of allergens, assessment of the potential allergenicity based on amino acid composition, assessment of the potential allergenicity based on dipeptide composition, and the evaluation of the similarity of protein sequence fragments to peptides representative of allergens. The results are verified on the basis of information found in the Allergome database, because it contains the results of experimental studies on allergens. Students can use text search options (*e.g.*, protein name, organism name from which the protein is derived) or a search based on the amino acid sequence using the "*Allergome Aligner*" option. Finding the compound in the Allergome database with the annotation that the allergen has been experimentally analyzed and registered by IUIS confirms the results of *in silico* prediction indicating that the specified protein is a potential allergen. Some of the algorithms offered by the AlgPred program can provide false negative results. The exercise, in addition to familiarizing students with examples of bioinformatic tools used in the field of immunology, also teaches criticism in relation to the obtained results. Excessive optimism in the interpreta-

tion of the results obtained using *in silico* methods is also perceived as a problem in scientific research [Boulesteix, 2010, 2015]. Discrepancies in the results of allergenicity predictions obtained using different algorithms appear also in scientific research [Gowthaman & Agrewala, 2008]. An example may be the results of the *in silico* analysis on the allergenicity of proteins transporting lipids published by Garino *et al.* [2016]. These authors used, among others, the AlgPred program, and the reliability assessment of the results was conducted using the amino acid sequences of known allergens.

The presented examples of practical use of the information contained in the internet repositories related to the broadly defined allergy are not exhaustive. The databases are constantly updated, which is associated with the deepening of knowledge in the allergy field. The presented examples of the use of *in silico* tools for analysis of the problems related to allergies suggest that research carried out this way can be used to assess the quality and safety of food. One of the advantages of databases is their commonness and accessibility also for people struggling with the symptoms of allergies. Thanks to the information contained in databases, users can expand their knowledge on the ingredients that should be eliminated from their diet [Sicherer *et al.*, 2015].

Scientific interest in broadly defined bioinformatic analysis inspires the development of new programs useful in knowledge extension, also on regarding allergies. An example is an increasing number of *in silico* tools updated and shared under the link: "*Immunology of proteins and peptides*" available in the "*Useful links*" tab of the BIOPEP database [Minkiewicz *et al.*, 2009] and in the OmicTools database (http://omictools.com/immunology-category) [Henry *et al.*, 2014].

CONCLUSIONS

Today, the databases and computer programs are often applied in the research on bioinformatics analysis of molecules. Bioinformatic tools are also useful in the field of allergology. The computer-aided methods and information provided in databases are an essential part of research on allergens and mainly their epitopes. Online databases provide an easier access to a range of information on the structure and properties of proteins and their fragments that may be responsible for the occurrence of hypersensitivity reactions. The use of internet data repositories may be suitable for developing new vaccines. This is related to the potential allergenicity of the molecule sequence analysis or epitope structure prediction. Determination of the structure of epitopes and their location in the peptide sequence of a protein is also important in food research. The identification of such fragments can enrich the knowledge about the immune-mediated hypersensitivity reactions and be useful to people tackling with this problem.

STUDY FUNDING

CONFLICT OF INTERESTS

Authors declare no conflict of interests.

REFERENCES

1. Ahsan N., Rao R.S.P., Gruppuso P.A., Ramratnam B., Salomon A.R., Targeted proteomics: Current status and future perspectives for quantification of food allergens. J. Proteom., 2016, 143, 15–23.
2. Ameratunga R., Crooks C., Simmons G., Woon S.T., Health risks and adverse reactions to functional foods. Crit. Rev. Food Sci. Nutr., 2016, 56, 318–325.
3. Arnon R., Van Regenmortel M.H.V., Structural basis of antigenic specificity and design of new vaccines. FASEB J., 1992, 6, 3265–3274.
4. Atwood T.K., Bongcam-Rudloff E., Brazas M.E., Corpas M., Gaudet P., Lewitter F., Mulder N., Palagi P.M., Schneider M.V., van Gelder C.W.G., GOBLET Consortium, GOBLET: The Global Organisation for Bioinformatics Learning. Education and Training. PLoS Comput. Biol., 2015, 11, Article No e1004143.
5. Bhasin M., Singh H., Raghava G.P.S., MHCBN: a comprehensive database of MHC binding and non-binding peptides. Bioinformatics, 2003, 19, 665–666.
6. Blythe M.J., Doytchinova I.A., Flower D.R., JenPep: a database of quantitative functional peptide data for immunology. Bioinformatics, 2002, 18, 434–439.
7. Borchers A., Teuber S.S., Keen C.L., Gershwin M.E., Food safety. Clin. Rev. Allergy Immunol., 2010, 39, 95–141.
8. Boulesteix A.L., Over-optimism in bioinformatics research. Bioinformatics, 2010, 26, 437–439.
9. Boulesteix A.L., Ten simple rules for reducing overoptimistic reporting in methodological computational research. PLoS Comput. Biol., 2015, 11, Article No e1004191.
10. Brazas M.D., Ouellette B.F.F., Continuing education workshops in bioinformatics positively impact research and careers. PLoS Comput. Biol., 2016, 12, Article No e1004916.
11. Breiteneder H., Chapman M.D., Allergen nomenclature, 2014, *in*: Allergens and Allergen Immunotherapy, 5th edition (eds. R.F. Lockey, D.K. Ledford). CRC Press: Boca Raton, FL, USA, pp. 37–49.
12. Brusic V., Information management for the study of allergies. Inflamm. Allergy – Drug Targets, 2006, 5, 35–42.
13. Brusic V., Millot M., Petrovsky N., Gendel S.M., Gigonzac O., Stelman S.J., Allergen databases. Allergy, 2003, 58, 1093–1100.
14. Bucholska J., Minkiewicz P., The use of peptide markers of carp and herring allergens as an example of detection of sequenced and non-sequenced proteins. Food Technol. Biotech., 2016, 54, 266–274.
15. Carrera M., Cañas B., Piñeiro C., Vázquez J., Gallardo J.M., Identification of commercial hake and grenadier species by proteomic analysis of the parvalbumin fraction. Proteomics, 2006, 6, 5278–5287.
16. de la Iglesia D., García-Remesal M., de la Calle G., Kulikowski C., Sanz F., Maojo V., The impact of computer science in molecular medicine: Enabling high-throughput research. Curr. Top. Med. Chem., 2013, 13, 526–575.
17. Di Costanzo M., Paparo L., Cosenza L., Di Scala C., Nocerino R., Aitoro R., Berni Canani R., Food allergies: novel mechanisms

and therapeutic perspectives. Methods Mol. Biol., 2016, 1371, 215–221.

18. Ding Y., Wang M., He Y., Ye A.Y., Yang X., Liu F., Meng Y., Gao G., Wei L., "Bioinformatics: introduction and methods", a bilingual Massive Open Online Course (MOOC) as a new example for global bioinformatics education. PLoS Comput. Biol., 2014, 10, Article No e1003955.

19. Dziuba M., Minkiewicz P., Dąbek M., Peptides, specific proteolysis products as molecular markers of allergenic proteins – *in silico* studies. Acta Sci. Polon. Technol. Aliment., 2013, 12, 101–112.

20. Erhardt G., Shuiep E.T.S., Lisson M., Weimann C., Wang Z., Zubeir I.E.Y.M., Pauciullo A., Alpha S1-casein polymorphisms in camel (*Camelus dromedarius*) and descriptions of biological active peptides and allergenic epitopes. Tropical Anim. Health Prod., 2016, 48, 879–887.

21. Ertl P., Molecular structure input on the web. J. Cheminform., 2010, 2, Article No 1.

22. Finn R.D., Mistry J., Tate J., Coggill P., Heger A., Pollington J.E., Gavin O.L., Gunesekaran P., Ceric G., Forslund K., Holm L., Sonnhammer E.L., Eddy S.R., Bateman A., The Pfam protein families database. Nucleic Acids Res., 2010, 38, D211-D222.

23. García B.E., Lizaso M.T., Cross-reactivity syndromes in food allergy. J. Investig. Allergol. Clin. Immunol., 2011, 21, 162–170.

24. Garino C., Coïsson J.D., Arlorio M., *In silico* allergenicity prediction of several lipid transfer proteins. Comput. Biol. Chem., 2016, 60, 32–42.

25. Gendel S.M., Allergen databases and allergen semantics. Regul. Toxicol. Pharm., 2009, 54, S7–S10.

26. Gołąb J., Jakóbisiak M., Lasek W., Stokłosa T. (Ed.), Immunology, 2010, PWN Scientific Publishing House, Warsaw, pp. 1–3, ISBN 978–83–01–15154–6 (in Polish).

27. Goodman R.E., Ebisawa M., Ferreira F., Sampson H.U., van Ree R., Vieths S., Baumert J.L., Bohle B., Lalithambika S., Wise J., Taylor S.L., AllergenOnline: A peer-reviewed, curated allergen database to access novel food proteins for potential cross-reactivity. Mol. Nutr. Food Res., 2016, 60, 1183–1198.

28. Goodman R.E., Wise J., Bioinformatic analysis of proteins in Golden Rice 2 to assess potential allergenic cross-reactivity. Food Allergy Research and Resource Program, University of Nebraska, expertise nr BIO-02–2006, 2006, 1–24. available on-line: [http://www.allergenonline.org/Golden%20Rice%202%20Bioinformatics%20FARRP%202006.pdf], December 2016.

29. Goodrow M.H., Harrison R.O., Hammock B.D., Hapten synthesis, antibody development and competitive inhibition enzyme immunoassay for s-triazine herbicides. J. Agric. Food Chem., 1990, 38, 990–996.

30. Gowthaman U., Agrewala J.N., *In Silico* tools for predicting peptides binding to HLA-class II molecules: more confusion than conclusion. J. Proteome Res., 2008, 7, 154–163.

31. Gűnther S., Hempel D., Dunkel M., Rother K., Preissner R., SuperHapten: a comprehensive database for small immunogenic compounds. Nucleic Acids Res., 2007, 35, D906-D910.

32. Harzing A.W., Alakangas S., Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. Scientometrics, 2016, 106, 787–804.

33. Hayes M., Rougé P., Barre A., Herouet-Guicheney C., Roggen E.L., *In silico* tools for exploring potential human allergy to proteins. Drug Discovery Today, Disease Models, 2015, 17–18, 3–11.

34. Henry V.J., Bandrowski A.E., Pepin A.S., Gonzalez B.J., Desfeux A., OMICtools: an informative directory for multi-omic data analysis. Database, 2014, Article No bau069.

35. Holton T.A., Vijayakumar V., Khaldi N., Bioinformatics: Current perspectives and future directions for food and nutritional research facilitated by a Food-Wiki database. Trends Food Sci. Technol., 2013, 34, 5–17.

36. Ivanciuc O., Schein C.H., Garcia T., Oezguen N., Negi S.S., Braun W., Structural analysis of linear and conformational epitopes of allergens. Regul. Toxicol. Pharm., 2009, 54 (3 Suppl.), S11–S19.

37. Jędrychowski L., Wróblewska B., Szymkiewicz A., State of the art on food allergens – a review. Pol. J. Food Nutr. Sci., 2008, 58, 165–175.

38. Jiménez-Saiz R., Benedé S., Molina E., López-Expósito I., Effect of processing technologies on the allergenicity of food products. Crit. Rev. Food Sci. Nutr., 2015, 55, 1902–1917.

39. Juhász A., Haraszi R., Maulis C., ProPepper: a curated database for identification and analysis of peptide and immune-responsive epitope composition of cereal grain protein families. Database, 2015, Article No bav100.

40. Kadam K., Sawant S., Jayaraman V.K., Kulkarni-Kale U., Databases and Algorithm in Allergen Informatics. Bioinformatics – Updated Features and Applications, Abdurakhmonov I. (Ed.), InTech, 2016, DOI: 10.5772/63083.

41. King T.P., Hoffmann D., Loewenstein H., Marsh D.G., Platts-Mills T.A.E., Thomas W., WHO/IUS Allergen Nomenclature Subcommittee, 1994, ACI News 6/2:38–44.

42. Koeberl M., Clarke D., Lopata A.L., Next generation of food allergen quantification using mass spectrometric systems. J. Proteome Res., 2014, 13, 3499–3509.

43. Lata S., Bhasin M., Raghava G.P.S., MHCBN 4.0, A database of MHC/TAP binding peptides and T-cell epitopes. BMC Res. Notes, 2009, 2, Article No 61.

44. Leung N.Y.H., Wai C.Y.Y., Shu S., Wang J., Kenny T.P., Chu K.H., Leung P.S.C., Current immunological and molecular biological perspectives on seafood allergy: a comprehensive review. Clin. Rev. Allerg. Immunol., 2014, 46, 180–197.

45. Liu Z.P., Wu L.Y., Wang Y., Zhang X.S., Chen L., Bridging protein local structures and protein functions. Amino Acids, 2008, 35, 627–650.

46. Mari A., Rasi C., Palazzo P., Scala E., Allergen databases: current status and perspectives. Curr. Allergy Asthma Rep., 2009, 9, 376–383.

47. 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.

48. Marti P., Truffer R., Stadler M.B., Keller-Gautschi E., Crameri R., Mari A., Schmid-Grendelmeier P., Miescher S.M., Stadler B.M., Vogel M., Allergen motifs and the prediction of allergenicity. Immunol. Lett., 2007, 109, 47–55.

49. McSparron H., Blythe M.J., Zygouri C., Doytchinova I.A., Flower D.R., JenPep: A novel computational information resource for immunobiology and vaccinology. J. Chem. Inf. Comput. Sci., 2003, 43, 1276–1287.

50. Minkiewicz P., Darewicz M., Iwaniak A., Bucholska J., Starowicz P., Czyrko E., Internet databases of the properties, enzymatic reactions, and metabolism of small molecules-search options

and applications in food science. Int. J. Mol. Sci., 2016, 17, Article No 2039.

51. Minkiewicz P., Darewicz M., Iwaniak A., Sokołowska J., Starowicz P., Bucholska J., Hrynkiewicz M., Common amino acid subsequences in a universal proteome-relevance for food science. Int. J. Mol. Sci., 2015, 16, 20748–20773.

52. Minkiewicz P., Dziuba J., Darewicz M., Iwaniak A., Michalska J., Online programs and databases of peptides and proteolytic enzymes – a brief update for 2007–2008. Food Technol. Biotech., 2009, 47, 345–355.

53. Minkiewicz P., Miciński J., Darewicz M., Bucholska J., Biological and chemical databases for research into the composition of animal source foods. Food Rev. Int., 2013, 29, 321–351.

54. Nakamura R., Teshima R., Talagi K., Sawada J.I., Development of Allergen Database for Food Safety (ADFS): an integrated database to search allergens and predict allergenicity. Bull. Nat. Inst. Health Sci., 2005, 123, 32–36.

55. Ortea I., O'Connor G., Maquet A., Review on proteomics for food authentication. J. Proteom., 2016, 147, 212–225.

56. Pearson W.R., Lipman D.J., Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci., 1988, 85, 2440–2448.

57. Pfiffner P., Truffer R., Matsson P., Rasi C., Mari A., Stadler B.M., Allergen cross reactions: a problem greater than ever thought? Allergy, 2010, 65, 1536–1544.

58. Radauer C., Nandy A., Ferreira F., Goodman R.E., Larsen J.N., Lidholm J., Pomés A., Raulf-Heimsoth M., Rozynek P., Thomas W.R., Breiteneder H., Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. Allergy, 2014, 69, 413–419.

59. Radauer C., Bublin M., Wagner S., Mari A., Breiteneder H., Allergens are distributed into few protein families and possess a restricted number of biochemical functions. J. Allergy Clin Immun., 2008, 121, 847–852.

60. Saeed H., Gagnon C., Cober E., Gleddie S., Using patient serum to epitope map soybean glycinins reveals common epitopes shared with many legumes and tree nuts. Mol. Immunol., 2016, 70, 125–133.

61. Saha S., Bhasin M., Raghava G.P.S., Bcipep: A database of B-cell epitopes. BMC Genomics, 2005, 6, Article No 79.

62. Saha S., Raghava G.P.S., AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res., 2006, 34, W202-W209.

63. Scalbert A., Andres-Lacueva C., Arita M., Kroom P., Manach C., Urpi-Sarda M., Wishart D., Databases on food phytochemicals and their health-promoting effects. J. Agric. Food Chem., 2011, 59, 4331–4348.

64. Sharma O.P., Das A.A., Krishna R., Suresh Kumar M., Mathur P.P., Structural Epitope Database (SEDB): A Web-based Database for the epitope, and its intermolecular interaction along with the tertiary structure information. J. Proteom. Bioinform., 2012, 5, 3, 084–089.

65. Shreder K., Synthetic haptens as probes of antibody response and immunorecognition. Methods, 2000, 20, 372–379.

66. Sicherer S.H., Leung D.Y.M., Advances in allergic skin disease, anaphylaxis, and hypersensitivity reactions to foods, drugs and insects in 2014. J Allergy Clin. Immunol., 2015, 135, 357–367.

67. Sicherer S.H., Epidemiology of food allergy. J. Allergy Clin. Immunol., 2011, 127, 594–602.

68. Singh M.K., Srivastava S., Raghava G.P.S., Varshney G.C., HaptenDB: a comprehensive database of haptens, carrier proteins and anti- hapten antibodies, Bioinformatics, 2006, 22, 253–255.

69. Singh K.V., Kaur J., Varshney G.C., Raje M., Suri C.R., Synthesis and characterisation of hapten–protein conjugates for antibody production against small molecules. Bioconjug. Chem., 2004, 15, 168–173.

70. Sircar G., Sarkar D., Bhattacharya S.G., Saha S., Allergen databases. Methods Mol. Biol., 2014, 1184, 165–181.

71. Siruguri V., Bharatraj D.K., Vankudavath R.N., Rao Mendu V.V., Gupta V., Goodman R.E., Evaluation of Bar, Barnase, and Barstar recombinant proteins expressed in genetically engineered *Brassica juncea* (Indian mustard) for potential risks of food allergy using bioinformatics and literature searches. Food Chem. Toxicol., 2015, 83, 93–102.

72. Tapal A., Vegarud G.E., Sreedhara A., Hegde P., Inamdar S., Kaul Tiku P., *In vitro* human gastro-intestinal enzyme digestibility of globulin isolate from oil palm (*Elaeis guineensis* var. *Tenera*) kernel meal and the bioactivity of the digest. RSC Adv., 2016, 6, 20219–20229.

73. The UniProt Consortium, UniProt: a hub for protein information. Nucl. Acids Res., 2015, 43, D204–D212.

74. Tong J.Ch., Song Ch.M., Tan P.T.J., Ren E.Ch., Sinha A.A., BEID: Database for sequence- structure-function information on antigen-antibody interactions. Bioinformation, 2008, 3, 58–60.

75. Trapp J., Web of Science, Scopus, and Google Scholar citation rates: a case study of medical physics and biomedical engineering: what gets cited and what doesn't? Australas. Phys. Eng. Sci. Med., 2016, 39, 817–823.

76. Vita R., Overton J.A., Greenbaum J.A., Ponomarenko J., Clark J.D., Cantrell J.R., Wheeler D.K., Gabbard J.L., Hix D., Sette A., Peters B., The immune epitope database (IEDB) 3.0. Nucleic Acids Res., 2015, 43, D405–D412.

77. Wróblewska B., Szymkiewicz A., Jędrychowski L., Impact of technological processes on changes in food allergies. Żywn. Nauk. Technol. Jakość, 2007, 6, 55, 7–19 (in Polish).