Role of Bioinformatics in Food Allergens: An Overview

Anoushka Kotra[1*], Nidhee Chaudhary[2] and Neetu Jabalia[3]

[1,2,3]Amity Institute of Biotechnology, Amity University, Uttar Pradesh, Noida
E-mail: [1]nushibushi20@gmail.com, [2]nchaudhary@amity.edu, [3]njabalia@amity.edu

Abstract—*Allergy is one of the most persistent health issue and the induction of recombinant amino acid polymers in food, has become a matter of public concern. It is a world-wide problem. Food allergenicity is referred as an untypical response towards the food by the immune system of the body. Allergy is steadily increasing health problems for all age groups in our society. These days bioinformatics is playing a vital role in developing improved technology for the detection and characterization of food allergens. Computational approaches are being increasingly utilized to evaluate the degree of similarity between a novel protein and known allergens within the context of a larger allergy safety assessment process. In this overview, we described the allergy databases and bioinformatics tools to identify the common proteins that may be at the root of multiple allergy syndromes.*

Keywords*: Allergy, Bioinformatics, recombinant*

1. INTRODUCTION

Food allergy is the response towards food that is not favored by the immune system of the body. Common foods that cause allergy are cow's milk, peanuts, eggs, shellfish, tree nuts, wheat, rice and fruit [2-4]. When IgE (Immunoglobin E) molecule binds to the food, the body produces an allergic response. This allergic response by the body can be witnessed through various signs and symptoms that may range from mild to severe. These include: vomiting, low bp (blood pressure), itchiness, swelling of tongue, hives, etc.

3 million children in the United States were known to have some type of food allergy [2].

There are mild to severe symptoms of food allergy, however all reactions won't be similar just because an initial reaction caused a few problems [12].

2. MECHANISM OF FOOD ALLERGY

The Food Allergy mechanism (Fig. 1) involves cytokines IL-4, IL-5 and IL-13 that are associated with the T-helper 2 response. The immune system of the body is biased towards this

Response and therefore induces IgE towards allergens via Bcells. This is followed by degranulation of mast cells when IgE receptors present on these cells are binded by the IgE. Thus, the body experiences itchiness, sneezing, etc when the mast cells release its mediators [5].

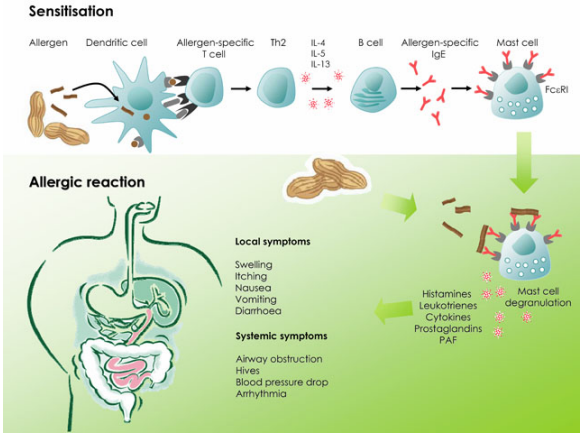

Fig. 1: Mechanism of food allergy.

Types of Food Allergy

According to FARE (Food Allergy Research and Education) there are eight foods that are responsible for 90% of food allergies (Figure2). They are:-

Fig. 2: Types of food allergies.

Milk allergy: Generally children are prone to this allergy. Proteins present in cow's milk such as casein or whey are

reacted upon by the body. Atopic diseases such as asthma, allergic rhinitis or eczema may also be found in people with this allergy.

Egg allergy: Children are the victims of this allergy, however the problem may resolve at a very young age. The people may also carry this allergy (Fig. 3) for their entire lives. A person may be prone to either the egg white or the yolk or both.

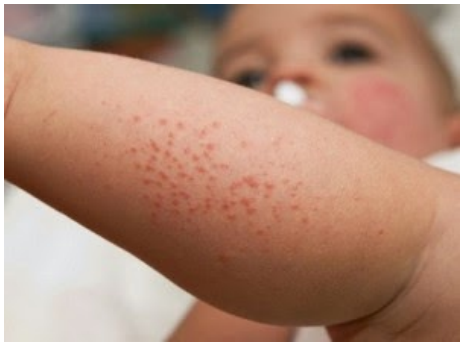

Fig. 3: Skin disease caused due to egg allergy.

Peanut allergy: This disorder persists for an entire life and is very serious. Anaphylaxis may occus. In severe cases a person may get restricted to breathing or it might lead to a cardiac arrest.

Other common allergies: Very little is known about the rest of them however, soy, wheat and tree nuts allergies are also considered to be life long disorders [6].

Bioinformatics

Bioinformatics is essential for the development of an improved technology that manages, detects and classifies food allergens via the application of computer science [1]. This information is represented in a statistical form. Allergen databases and data sources are all-purpose biology related databases that consists of annoted entries of biological sequences. Recently conducted studies show common molecular features of proteins from various sources, which account for clinically significant cross reactivity and sensitivity. For instance, proteins similar to the allergenic proteins found in peanuts that were extracted and their sequences reacted with IgE, were later found in other foods such as tree nuts, soy and legumes. It is important to know that incomplete information to eradicate the probability that there isn't any sequence similarity between a protein and a known allergen can cause an allergic reaction. Therefore, in such a situation one could rely on the computational approach that is the bioinformatics tool in order to compare these known allergens [10].

Bioinformatics is used for the management of biological information via application of computer science. The discovery of the sequence of insulin in 1950's by Frederick Sanger made the use of computers essential in molecular biology. To search sequences from more than 260,000 organisms, containing over 190 billion nucleotides, computer programs such as BLAST are used on a daily basis [7]. Apart from sequence analysis it also helps in predecting genes (genome annotation). The first genome annotation software was designed by Owen White in the year 1995, who was part of the team at The Institute of Genomic Research. He sequenced and analysed the first genome of a free-living organism to be decoded, the bacterium was *Haemophilus influenzae* [8]. Sequences of known allergens and B-cell epitopes are all included in a ADFS that is an allergy based database [11].

The aim of Human Genome Project is determination of the sequence of the entire human genome approximately three billion base pairs will be reached by the year 2002. The basic concept of bioinformatics is the modification of molecular biology with computer science. It helps in understanding human diseases. Drug discovery can be achieved by identifying new molecular targets of drug [13].

Common uses of bioinformatics include, the identification of candidate genes and nucleotides(SNPs). Bioinformatics has become an important part of many areas of biology. For example, in experimental molecular biology, bioinformatics techniques such as image and signal processing allow extraction of useful results from large amount of data. In genetics and genomics field, it helps in sequencing of genomes. It even plays an important role in the analysis of gene and protein expression and regulation. Historically, the term bioinformatics did not mean what it means today [14-16]. First protein compilation and sequencing of of databases was done by Dayhoff [17]. It showed new paths for the sequence alignment and evolution [18].

3. ALLERGEN DATABASES AND TOOLS

Genbank, EMBL, DDBJ, PIR, SWISS-PROT and PDB are the popularly used public databases. The main aim of these databases is to collect, annotate and provide access to enteries of sequences. SWISS-PROT and PIR are manually anootated protein databases. They have a higher detailed annotation than GenPept, trEMBL and DAD. The identification of cross reacting allergens is done by cross-indexed databases. Allergome is a comprehensive database that carries information about IUIS and non-IUIS recognized allergens.

IUIS: The IUIS (International Union of Immunological Socities) allergen nomenclature sub-committee under the auspices of WHO(World Health Organization) and IUIS, maintains systematic nomenclature of allergic proteins and publishes a database of approved allergen names on its website, www.allergen.org [9].

IFBC/ILSI: The IFBC/ILSI is expanded as International Food Biotechnology Council and International Life Science Institute.Sponsorship for assessment of transgenic proteins in crop plants was provided by the IFBL. A list of protein allergen sequences is submitted to public databases.

FARRP: The Food Allergy Research and Resource Programme (FARRP) database consists of allergens that contain: species of origin, common and nomenclature names and accession no. of the source(linked to ENTREZ).

ALLALLERGY: This database (Figure4) has more than 4500 enteries. It includes chemical allergens that contain only proteins. Data is regularly updated and information is fully refrenced.

Fig. 4: Allallergy database.

SDAP: SDAP (Structural Database of Allergic Proteins) is a web server that provides rapid, cross-referenced access to the sequences, structures and IgE epitopes of allergenic proteins [9]. Structure of SDAP involves website (http://fermi.utmb.edu/SDAP/), opening the main search page and selecting an allergen of interest. The descriptive page would contain a summary of all data including official name, source, type, brief description and aession no's.from SWISS-PROT, NCBI, etc. The information is cross-refrenced to original data sources. One can access them by clicking on links from each allergen page.

4. APPLICATION OF BIOINFORMATICS IN FOOD ALLERGY

With increase in information about sequence, structure and IgE epitopes of allergens, the bioinformatics search tools allow users to access the information through various databases. There are various methods for comparing sequences of allergens provided by SDAP. BLAST and FASTA databases are used in determining the amino acid sequence of allergens similar to the other sequences. SDAP provides the Pfam grouping for most of the allergens. It rapidly compares the resemblance of allergens with protein. Names of newly identified allergens can also be determined using SDAP.

Modern bioinformatics algorithms compare local sequences and global predictions of a structure to find similarity. FASTA algoithm is used to find local high scoring alignments between a protein or neucleotide pair. BLASTp is the basic local alignment search tool for proteins.The above mentioned both the programs can be affectively used to predict similarity.These bioinformatics algorithms is based on the principle that if two proteins share enough linear sequence then they will also share three dimensional structure. Hence, they will even have similar functional homology. Homologous proteins share allergenic cross-reactive confirmational and linear epitopes in comparison to unrelated proteins. This is due to the similarity in sequence and structure. Homologous proteins share only 25% amino acid identity in case of 200 amino acid overlapping. This is not generally enough to indicate IgE related cross-reactivity. On the other hand allergenic cross-reactivity caused by protein that share confirmational or linear epitopes is rarely 50% identity and requires 170% amino acid identity. Such high levels of identity are detected using FASTA or BLASTp. These databases can be used for the following purpose:-

Comparison of sequence of allergic proteins: The sequences of allergic proteins can be compared using this database. The first step involves the determination of potentially cross reactive proteins by determining the sequence similarity to other allergens. To measure the number of matches with the same sequence similarity provides an E-value or Expectation value. A low E-value means that there is high probability of sequence match.

Predicting Allerginicity: The term "motif" refers to residue conservation in discrete areas of related allergenic proteins of clinical cross-reactivity as in case of IgE binding. A group of cross-reactive allergenic proteins, the IgE epitopes areas will have common physical chemical properties (PCPs). In this method first we align the sequences of known allergens that are related to one another. For eg- tropomycin or vicilin family. The Physical Chemical Property based Motif analyser (PCPMer) finds sequence motifs in protein families by identifying regions with highly conserved PCPs.

Sequence combination and structural information: The 3-Dimensional view of the IgE binding sites can provide missing information about the relationship between structure and sequence. A program named GETAREA helps in determining the residues that are on the surface of an allergen. This program is even indulged with the site where the data can be quickly accessed for SDAP allergens. SDAP allows direct access to the experimental structures conserve residues on the surface of proteins for the detection of common areas [10].

5. CONCLUSION

Proteins are known as allergens because it enhances responses in patients. Basically, allergens are the advanced form of other proteins with similar surface areas. During development of disease it is possible that they have been the true responsive

antigens. From the recent studies it has been found that the sequence and structure of allergenic proteins from pollens and food has revealed about allerginicity and cross-reactivity.

Some of the tools of SDAP can help in finding the structural and functional relationship among allergens. It even helps in identifying cross-reacting antigens. These methods provide indication that certain proteins may be cross-reactive. Thus, these predictions help in developing guidelines of diets for individual patients and in the development of desgning specific immunotherapy.

Allergy related databases is used in collecting accessing and using data. In terms of sophistication of bioinformatics, a database shoud firstly comply with the official classification of allergens. And finally, should comprise non-redundant enteries of allergens with their features such as biochemical, structural, functional and clinical features. They are not restricted to the sequences, structures, sources, reference and relevant links.

The allergy related data is reaching new heights due to the advancement in genomic,proteomic and molecular biology techniques.

6. ACKNOWLEDGEMENT

We are grateful to our Director of Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida for his constent support and encouragement during this study.

REFERENCES

[1] Gendel,“Bioinformatics and Food Allergens”,2004.

[2] Anthony, S. Fauci, “National Institute of Allergy and Infectious diseases”, Food Allergy an Overwiew, July 2012.

[3] Sicherer, S. H. and Sampson, H. A., “Food Allergy:Epidemiology,pathogenisis,diagnosis, and treatment”, Allergy Clin Immunol,133,2,February 2014,pp.291-307.

[4] Nowak-Wegrzyn, Katz, Mehr and Koletzko, “Non-IgE-mediated gastrointestinal food allerg”, Journal of Allergy and Clinical Immunology, 135,5,May 2015,pp.24-1114.

[5] http://ucfa.nl/food-allergy/mechanisms/ Utercht Center for Food Allergy,2015.

[6] http://www.healthline.com/health/allergies/common-food-allergies#Overwiew1

[7] Benson, D. A., Karsch, Mizrachi, I., Lipman, D. J., Ostelle J., Wheeler D.L., “GenBank”, Nucleic Acids Res.,36,January 2008.

[8] Flieschmann, R. D, Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A. and Merrick, J. M., “Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.”,Science, 269, 5223,July 1995,pp.469-512.

[9] Radauer, C., Nandy, A., Ferriera, F., Goodman, R. E., Larsen, J. N., Lidholm, J., Pomes, A., Raulf-Heimsoth, M., Rozynek, P., Thomas, W. R., and Breiteneder, H., “Update of WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences”, Allergy,69,2014,pp.413-419.

[10] Ivanciuc O., Schein C.H. and Braun W., “SDAP database and computational tools for allergenic proteins”, Nucleic Acids Res,1,January 2003,pp.359-362.

[11] Nakamura R. and T.R, “Development of Allergen Database for Food Safety (ADFS)-an integrated database to search allergens and predict allergens,2005.

[12] http://acaai.org/allergiess/types/food-allergies

[13] http://www.bioplanet.com/what-is-bioinformatics/

[14] Hogeweg P. and David B., “The roots of Bioinformatics in Theoretical Biology”, Computational Biology, 7,3,2011

[15] Hesper B. and Hogeweg P., “Bioinformatica-een werkconcept”, kameleon,1,6, 1970,pp.28-29

[16] Hogeweg P., “Simulating the growth of cellular forms”,Simulation, 31,3,1978,pp.90-96

[17] Dayhoff and M.O., “Atlas of protein sequence and structure,National Biomedical Research Foundation”,1996,pp.215

[18] Eck R.V.,Dayhoff and M.O., “Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences”,Science,152,3720,1996,pp.363-366.