Identification and Structural Insights of Allergens in Anacardium occidentale

A dissertation

Submitted in partial fulfilment of the requirement

For the award of the degree of

Masters of Technology

In

Biotechnology

(July 2023)

Under the guidance of

Dr. Atul Kumar Upadhyay
Assistance Professor
Dept. of Biotechnology
TIET, Patiala

Submitted by

Yadvi

Roll No. 602104019

DEPARTMENT OF BIOTECHNOLOGY
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY,
PATIALA (PUNJAB) -147004

DECLARATION

I, Yadvi, hereby declare that the project work entitled **"Identification and Structural Insights of Allergens in *Anacardium occidentale* "** is an authentic record of my own work carried out at, Thapar Institute of Engineering and Technology, Patiala as the requirement of 12-month dissertation for the award of the degree of Masters of Technology in Biotechnology, under the guidance of Dr. Atul Kumar Upadhyay, during July 2022 to July 2023. Previously no other Institute or University has ever submitted this research work as a project or thesis to receive another degree or diploma.

Yadvi
602104019
M. Tech. Biotechnology
DBT, TIET
Patiala- 147004

Place: Patiala
Date: 14th July 2023

CERTIFICATE

This is to certify that the dissertation titled, "**Identification and Structural Insights of Allergens in *Anacardium occidentale***" submitted for the partial fulfillment of the requirements for the award of the Masters in Technology in the Department of Biotechnology at Thapar Institute of Engineering and Technology, Patiala is an authentic work carried out by **Yadvi, 602104019**, under my supervision and guidance.

To the best of our knowledge, the dissertation work embodied in this thesis has not been submitted to any other University/ Institute for the award of any Degree or Diploma.

(Dr. Atul Kumar Upadhyay)
Assistant Professor
Department of Biotechnology
Thapar Institute of Engineering and Technology
Patiala- 147004

ACKNOWLEDGEMENT

*First and foremost, I express my deepest gratitude to **'Lord Shri Krishna Ji'** for being with me from the beginning, till the end.*

*I would like to express my profound gratitude and deep regards to my guide or advisor **Dr. Atul Kumar Upadhyay,** Assistant Professor, Department of Biotechnology, TIET for his exemplary guidance, monitoring, and constant encouragement throughout the degree. The guidance and help given by him from time to time shall carry me a long way in the journey of life on which I am about to embark.*

*I am also thankful to **Prof. Padmakumar Nair,** Director, Thapar Institute of Engineering and Technology(TIET), and **Dr. MS Reddy**, HOD, Department of Biotechnology, Thapar Institute of Engineering and Technology(TIET) for their priceless and inexpressible help and encouragement during the course of my research work.*

Indeed, my words are not sufficient to elucidate my profound sense of reverence and in debt to my whole professors for the encouragement and immense knowledge, creative suggestion, and moral support during the degree.

I acknowledge the guidance and help extruded to Research Scholars Ms. Rashi Chugh, and Ms. Tanmayee Basu and my friends for memorable, sweet, and thoughtful company.

*My vocabulary utterly fails in expressing my accolade to my **family** for believing in me and for their innumerable selfless sacrifices, silent prayers, encouragement, and unconditional love throughout my life. They have always been there for me, whenever I need them.*

Place: Patiala Yadvi

Date: 14th July 2023 602104019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF GRAPHS

LIST OF ABBREVIATIONS

Abbreviations	Full form
°C	Degree Celsius
%	Percent
>	Greater Than
1/3rd	One Third
3D	Three Dimensional
µl	Microliter
g	Gram
µg	Microgram
kD	Kilodalton
ml	Milliliter
sec	Seconds
ACC	Auto Cross Covariance
ARP	Antigen Representing Peptides
AllFam	Allergen Families
BLAST	Basic Local Alignment Search Tool
BLOSUM 50	Blocks Substitution Matrix -50
DBPCFC	Double-Blind Placebo-Controlled Food Challenge
DNA	Deoxyribonucleic Acid
ELISA	Enzyme-Linked Immunosorbent Assay
E-value	Expect Value
FAO	Food And Agriculture Organization
FARRP	Food Allergy Research and Resource Program
FASTS	Fast Alignment
FDA	Food And Drug Administration
GO	Gene Ontology
HAMMER	Hidden Markov Model Method
IBS	Irritable Bowel Syndrome
IEBD	Immune Epitope Database and Analysis Resource
IFBC	International Food Biotechnology Council
ILSI	International Life Sciences Institute

I-TASSER	Iterative Threading Assembly Refinement
LC-HRMS	Liquid Chromatography- high Resolutions Mass Spectrometry
LC-MS	Liquid Chromatography – Mass Spectrometry
LFD	Lateral Flow Devices
LOD	Limit of Detection
LOQ	Limit of Quantitative
MUSCLE	Multiple Sequence Comparison by Log Expectation
NCBI	National Center for Biotechnology Information
NIAID	National Institute of Allergy and Infectious Diseases
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
Pfam	Protein Families Database
PSIPRED	PSI-Blast Based Secondary Structure Prediction
RMSD	Root Mean Square Deviation
SAXS	Small-Angle X-Ray Scattering
SDAP	Structural Database of Allergenic Proteins
SVM	Support Vector Machine
T value	Tanimoto Coefficients
UHPLC-MS/MS	Ultra- High Performance Liquid Chromatography – Mass Spectrometry
WEGO	Web Gene Ontology Annotation Plot Server
WHO/IUIS	World Health Organization and International Union of Immunological Societies

OBJECTIVES

1. Screening and Identification of Putative allergens in *Anacardium occidentale*.
2. Sequence-based Characterization of predicted putative allergens of *Anacardium occidentale*.
3. *In silico* validation of allergenicity among predicted allergens of *Anacardium occidentale*.

ABSTRACT

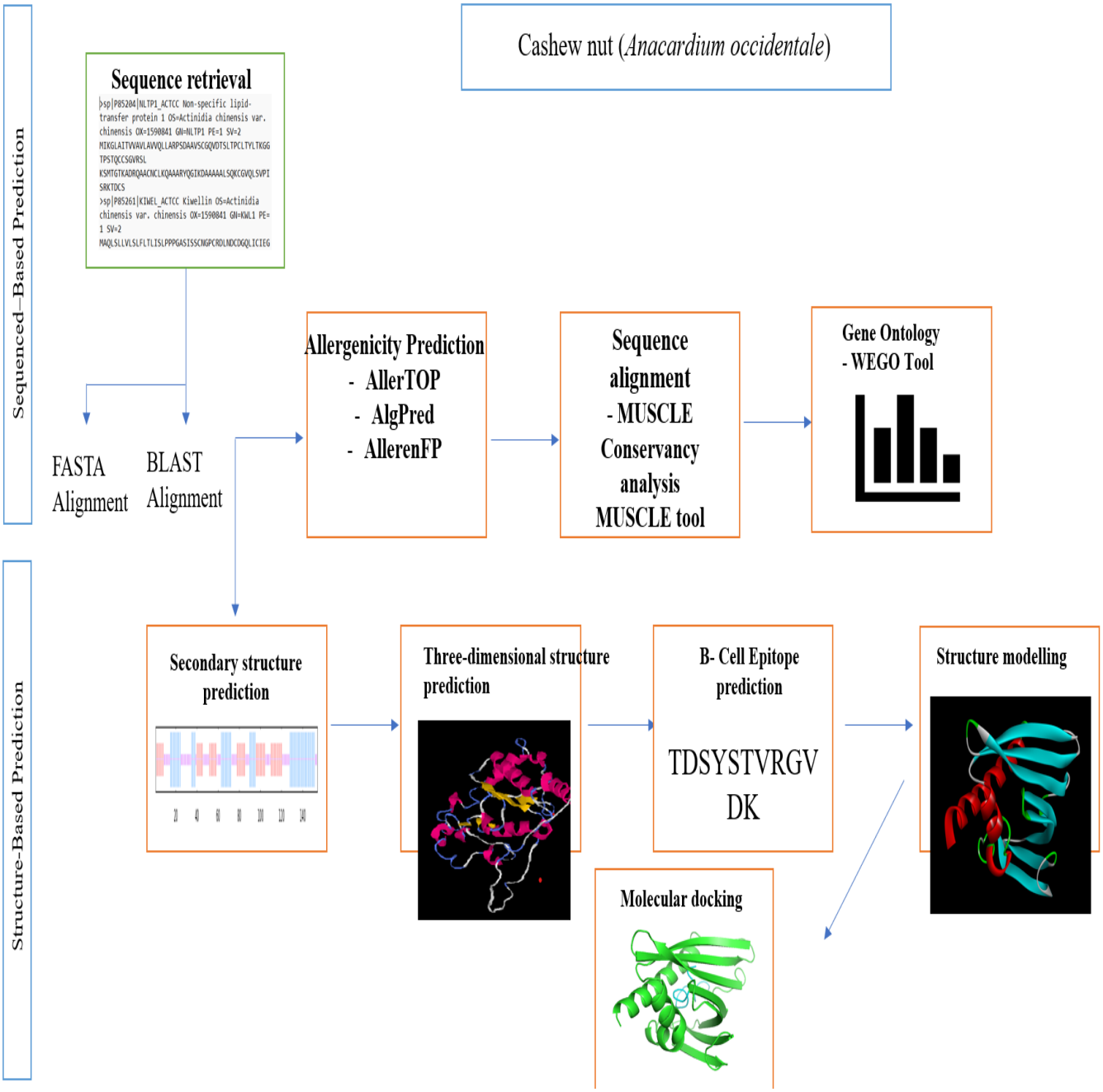
BACKGROUND: Food allergies are allergic reactions that are triggered by food ingredients and affect the immune system, they are most frequently caused by plant-derived foods, especially in adults and children. For tracing the allergens in food there have been developed several methods including both computational and conventional methods, but for preliminary assessment of food allergenicity, the use of computational tools may be more advantageous than using conventional methods. In this study, in silico tools were used to evaluate and validate the allergenic potential of cashew protein.

RESULTS: The cross-reactivity of cashew proteins with food allergens was evaluated using the Fast Alignment (FASTA) and Basic Local Alignment Search Tool (BLAST) algorithm-based sequence alignment. Eleven cashew proteins were cross-reactive with known food allergens by consensus approach using both FASTA and BLAST algorithm-based sequence alignment. BLAST data shows the E-value and the percent identity and FASTA alignment demonstrate >50% sequence identity of eleven cashew proteins. AllergenFP, AlgPred, and Allermatch – allergenicity predicted software predicted that eight out of eleven cashew proteins were potential allergens on the basis of their physicochemical properties. According to the sequence alignment using the MUSCLE tool, the cashew protein, and known food allergens were found to have 30.4%-66.8% conservancy. amino acid comparison and secondary structure between cashew protein and known food allergens were predicted and compared using the PHD fold server and PSIPRED. For quality assessment, three-dimensional structure, and superimposition of 8 cashew proteins with food allergens were generated. A protein family analysis was determined using Pfam, conserved domain databases, HMMER, and InterPro databases. Based on the GO accession number, the WEGO tool was used to visualize the gene ontology data. SWISS-MODEL server is used to model the structures of selected cashew proteins. B-cell epitopes for the selected cashew proteins were predicted and their structure modeling for the epitope was conducted by PepFold 3.5 server and subjected to docking in ClusPro 2.0 server. Several docked models were produced through docking; however, the lowest binding energy model was selected for further evaluation.

CONCLUSION: Preliminary information on the cross-reactivity and potential allergenicity of cashew proteins is provided by In silico technologies. The prelusive allergenicity of allergen sources can be assessed using this method.

Keywords: cashew allergy; food allergy; food allergen; allergenicity assessment; in silico allergenicity assessment; cross-reactivity.

GRAPHICAL ABSTRACT



CHAPTER 1 INTRODUCTION

Allergies are now recognized as a matter of public health. Allergies occur in the human body whenever the immune system reacts to a foreign substance. According to the Expert Panel Report published in 2010 and sponsored by the National Institute of Allergy and Infectious Diseases (NIAID) (Boyce et al., 2010). Food allergy is defined as "an adverse health effect emerging from a specific immune reaction that takes place reproducibly on exposure to a given food" (Burks et al., 2012). Whereas food intolerance is defined as "nonimmune responses which include metabolic, toxic, pharmacologic, and undefined mechanisms" (Gargano et al.,2021). The most commonly discovered food allergens are found naturally in seeds, nuts, and cereals, particularly peanuts. Fresh fruits and vegetables have also been connected to allergy reactions of varied degrees of severity, which can be fatal. Several important types of allergens in fruits, tree nuts, vegetables, legumes, and, cereals, such as the nonspecific lipid transfer proteins, the 2S albumin seed storage proteins, and the cereal a-amylase and protease inhibitors. (Breiteneder et al.,2004). Food reactions can have negative effects and can cause both immune-mediated and non-immune-mediated intolerances (Turnbull et al., 2015; Afzaal et al., 2023). Food allergies can cause a variety of symptoms, including coughing, nausea, vomiting, pharyngitis, and abdominal pain (Lee et al., 2023).

Food allergy has been a major concern for patients as well as physicians worldwide, owing to its rising prevalence and increased healthcare utilization. People of all ages have been reported to be impacted by food allergies (Höfer et al., 2023). According to previously reported cases analysis, 90% of food-allergic reactions in children are because of items such as milk, egg, peanut, soy, and wheat and 85% of the food-allergic reaction are due to consumption of shellfish, tree nuts, fish, and peanuts in youngsters (Lee et al., 2006; Hefle et al., 1996). An allergic person can react to several allergens, and typically, multiple allergens can be found in a single allergenic source. The ideal way to determine food allergens must therefore accomplish two objectives: first, to minimize the potential risk of allergy, allergenic components in food must be fully identified and secondly, to reduce its impact on quality of life, non-allergenic ingredients must then be completely excluded. (Matricardi et al., 2016). Most of the time allergens are glycoproteins in nature which range in size from 10- 70 kd, which are responsible for triggering immune response in the body.

Food allergens can sensitize the host in the gastrointestinal tract because it is resistant to acid, heat, and proteolysis. The majority of IgE-mediated food allergies are caused due to the foods such as gluten-containing cereals, eggs, milk, crustacean shellfish, sesame, fish, peanuts, and specific tree nuts (cashew, hazelnut, pecan, walnut, almond, and pistachio) (Nurmatov et al., 2017; Elghoudi et al.,2022). As a result, these items are frequently referred to as the "BIG 8" . According to the United Nations' Food and Agriculture Organization (FAO) (FAO, 1995) (Metcalfe et al., 1996). Sesame has been added to the "Big 8" as a replacement for soybeans, which have been removed from the list of the world's top eight allergens (FAO 2020). (Chen et al., 2022). Around 5% of young children and 3-4% of people in developed nations suffer from food allergies. Furthermore, peanut and tree nut allergies cause long-term allergy-related problems. Tree nut is the collaborative term that is used to describe the nuts which grow on the tree (Wang et al., 2020). Numerous tree nuts can cause an IgE-mediated food allergy reaction, like almond (*Prunus dulcis or Amygdalus communis L.*), brazil nut (*Bertholletia excelsa*), cashew nut (*Anacardium occidentale*), hazelnut (*Corylus avellana*), macadamia nuts (*Macadamia ternifolia*), pecan nuts (*Carya illinoiesis*), pine nuts (*Pinus pinea* and other *Pinus* species), pistachio nut (*Pistacia vera*) and walnut (*Juglans regia*) (Costa et al., 2016), Although they are not botanically related, peanut and tree nut allergies have numerous clinical similarities. Two of the most often reported foods to trigger IgE-mediated food allergy responses are peanuts and tree nuts. These allergic reactions to nuts can have serious and even fatal consequences (Crespo et al., 2006). Tree nut allergy is reportedly seen in up to 0.05% to 4.9% of individuals worldwide, according to research from a systematic review. (McWilliam et al., 2015). Among all tree nuts, cashews are the third most produced in the world. Cashew nuts rank third in worldwide production among all tree nuts (Rico et al., 2016). The cashew (*Anacardium occidentale*), which includes 9 species of the genus Anacardium, is a member of the Anacardiaceae family (Cíntia et al., 2019). Due to their high level of lipids, fiber, sterols, vitamins, amino acids, and minerals, cashews are an extremely nutrient-dense food (Ros et al., 2010), like other nuts, cashew nuts can also induce allergies, and the severity of cashew nut allergies might vary from one region to the next (Costa et al., 2016). There are officially recognized cashew allergens named Ana o 1, a vicilin protein, Ana o 2, a legumin protein, and Ana o 3, a 2S albumin, all of which have been listed in the World Health Organization and International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature database (Fang Wang et al., 2002; Fang Wang et al., 2003; Robotham et al.,2005). These three allergens from cashews, Ana o 1, Ana o 2, and Ana o 3, are all

categorized as seed storage proteins. Other tree nuts, legumes, and seeds are known to contain the same allergens from these families of seed storage proteins (Borres et al., 2022).

For the identification of allergens in food, several approaches are available which include both conventional and computational methods (Singh, A et al., 2020). Hence, to study food allergenicity, computational tools can be used quickly and have advantages over the conventional method because of many limitations like low sensitivity of results, and collection of blood samples from suspectable hosts for detection of Ig-E binding proteins. The majority of computational tools have open access and user-friendly interfaces that assist in producing accurate data that can be used as early indicators of allergenic potential. Over the past decade, lots of allergens have been characterized and easily accessible through different online databases, such as (WHO/IUIS) Allergen Nomenclature Database (www.allergen.org), and (FARRP) The food allergy Research and Resource Programm (www.allergenonline.org) which is an allergen protein database (Vashisht et al., 2023). In silico methods accounts for both sequence-based and structure-based approach for determining the allergenicity of a protein. By analyzing physicochemical properties, the potential cross-reactivity of putative allergens with known allergens can be discovered. The main aim of this study is to identify and validate the likelihood of cashew proteins being an allergen.

2.1 Food Allergy

Van Pirquet in 1906, first give the term “Allergy”. Allergies occur in the human body whenever the immune system reacts to a foreign substance. Food allergy is defined as "an adverse health effect emerging from a specific immune reaction that takes place reproducibly on exposure to a given food" (Burks et al., 2012). Based on the particular symptoms and the sensitization to the specific food, food allergy is diagnosed. Food allergy immunological responses can be categorized as IgE-mediated, non-IgE-mediated, or a combination of both. Sensitization occurs as the first sign for the IgE mediated food allergy, secondly, symptoms develop on exposure to a particular food. However, when T-cell mediated become predominant, it causes non-IgE-mediated food allergy and it may have historical evidence in the past (Borres et al., 2022).

2.2 Food Intolerances

When Undesirable reactions occur due to a specific food are terms as food intolerance. These are mainly non-allergic food reactions but show some symptoms and they don’t involve the immune system. Food intolerance is defined as "nonimmune responses which include metabolic, toxic, pharmacologic, and undefined mechanisms" (Gargano et al., 2021). However, sometimes organic pathophysiological processes involve food intolerance like lactose intolerance which occurs due to the enzyme deficiency that breakdowns to the lactose. There are some cases like irritable bowel syndrome (IBS) patients, which don’t explain the main reason for the food intolerance (Borres et al., 2022).

2.3 Tree Nuts Allergy

Tree nuts are a common source of food allergy that induces the IgE- mediated allergic reactions (Nagakura et al., 2020). Tree nut allergic reactions affect human health and can be serious and life-threatening. According to a systematic review, tree nut allergy is prevalent in up to 0.05% to 4.9% of people globally (McWilliam et al., 2015). Mainly, the tree nuts grow on the trees and are the fruit that has a hard shell protecting the kernel. Depending upon the protein sequences of the tree nuts they may show high similarity but some are distantly related and can be cross-reactive (Borres et al., 2022). Due to the high amount of seed storage protein, vitamins,

minerals, Fiber, carbohydrates, and many more, tree nuts contain healthy nutritional values. Johnson et al in 2014 identified the increase in anaphylaxis cases because of tree nut allergies, and find out that 373.3% of cases of anaphylaxis increased in its 10-year study period, also noted that the cashew allergy increased while the other tree nuts allergy remail stable during its study period (Borres et al., 2022).

Sicherer et al., 2010 revealed that the self-reported tree nuts allergy prevalence increased from 0.6% to 2.1% over the past 11 years in children, while in adults it remain constant. Protein superfamilies involve in tree nuts allergies are 11S globulins (vicilin), 2S albumins, and legumins having different biological functions. For diagnostics of tree nut allergies, Component-resolved diagnostics (CRD) are used and show accurate result trough measuring the s-IgE to proteins in place of measuring it with whole allergens (Blazowski et al., 2019; Datema et al.,2018).

2.4 Mechanism of IgE-mediated food allergy

An allergenic reaction takes place when the body is subsequently exposed to the allergen protein. After the exposure, the allergen protein binds and cross-links to IgE antibodies on the mast cell/basophil, leading to the degranulation of histamine along with some other inflammatory mediators (Wangorsch *et al* 2009).

Normally, the immune system raises immune responses to protect the body from foreign invaders, such as bacteria or viruses. A person having an allergy, their immune system also reacts to substances producing allergic reaction such as substances called allergens can come from the patient's natural environment, food, medication, or latex products.

Most allergies are mediated by a class of antibodies called IgE (Immunoglobulin E). IgE is produced when the body is first exposed to an allergen, production of IgE is activated by a subtype of T lymphocytes known as type 2 helper T cells, IgE molecules then bind to their receptors on the surface of mast cells and basophils. The first exposer is usually asymptomatic but the body is now sensitized upon re-exposer to the same antigen, the antigen bind to adjacent IgE molecule, and brings their receptors together, triggering a signaling cascade that induces the release of histamine and other inflammatory chemicals (Wangorsch et al., 2009).

These chemicals cause dilation and increase the permeability of blood vessels, and mucus secretion stimulation of sensory nerve and are responsible for allergic symptoms which can range from mild to severe. The reaction is immediate within minutes of contact with an allergen.

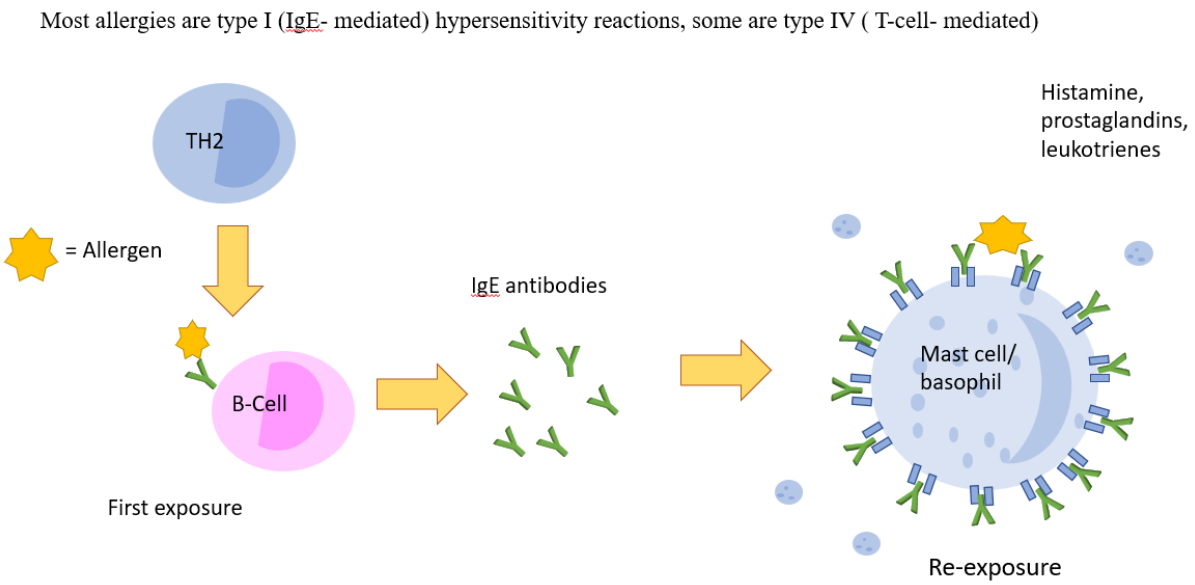


Figure 1: Mechanism of IgE-mediated food allergy

2.5 Big 8 Allergens

A paper about food allergy was published in 1993 and was presented by the Codex Committee on Food Labeling, which was followed by a Food and Agriculture Organization (FAO) Technical Consultation on Food Allergens in 1995. Due to this concern about food allergy come and in 1999, the Codex General Standard for the Labeling of Prepackaged Food, adopted the list of food allergens (Messina et al., 2020). Mainly there are eight allergens which are known as the “Big 8” allergens by the Food and Agriculture Organization (FAO). Congress adopted this list of Big 8 allergens and spread awareness about these allergens which was established by Codex in 1999. The paper published in 2004 by the Congress associated related to the establishment of the Big 8 allergens does not provide the data, but after one year, US Food and Drug Administration (FDA) reported the specific data for the Big 8 allergens in the report entitled “Approaches to Establish Thresholds for Major Food Allergens and for Gluten in Food” (Messina et al., 2020). The big 8 allergen includes Milk, eggs, wheat, peanuts, tree nuts, soybean, fish, and crustacean shellfish according to the United Nations' Food and Agriculture Organization (FAO) (FAO, 1995). Sesame has been added to the "Big Eight" as a

replacement for soybeans, which have been removed from the list of the world's top eight allergens (FAO 2020). (Chen, H et al., 2020).

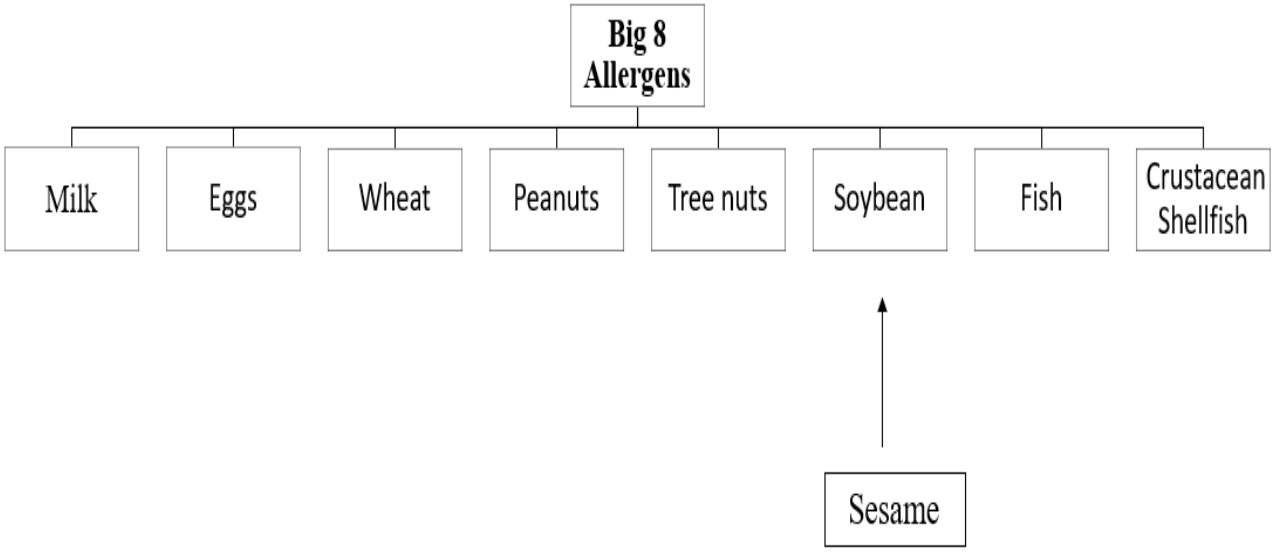


Figure 2: Flowchart illustrating the Big 8 Allergens

2.6 Cashew

Cashew scientifically known as the *Anacardium occidentale* belongs to the *Anacardiaceae* family and includes around 9 species of the *Anacardium* genus (Mendes et al., 2019). The cashew tree mainly comes in the category of the tropical evergreen perennial tree which is commonly recognized for its seed and fruits (*Borres et al., 2022)*. The fruit of the cashew tree well known as the cashew apple is a pseudo fruit that has a sugary and sweet flavor having a sweet accent. Cashew apple is used to prepare jams, juices and alcoholic beverages which are from its juicy and pulpy part and this part of cashew is very popular *(Mendes et al., 2019)*. Similarly, a seed i.e. actually a cashew nut, grows on the hard shell at the base of the peduncle and is rich in triglycerides, proteins, and carbohydrates. Cashew nuts are a popular snack all around the world and are usually eaten roasted and salted.

Cashew (*Anacardium occidentale*) is native to South America i.e. Brazil, but they are now produced in several countries in Asia and Africa according to FAOSTAT, 2016. Cashews in

terms of consumption rank third and terms of trade come in the second position after almonds (Mah et al., 2017). They contain vitamins, amino acids, fats, sterols, fibres, and minerals due to which they have high nutritional values, but their medical efficiency of it is not yet clinically studied (Ros, 2010). Although high consumption of cashew can lead to heart disease because of the fatty acids present in them (Mah et al., 2017). But they are used as traditional medicine for curing inflammation, skin infections, asthma, diabetes and many more (Alasalvar et al., 2008). Cashews may cause allergies to some people mainly adults and children, but their consumption is increasing in the past few years.

2.7 Cashew allergens

The two superfamilies of plant proteins that contain allergens most frequently found are cupin and prolamin. Both of these protein families function as the plant's defence mechanism. The legumin and vicilin types of allergenic seed storage proteins, which are found in peanuts, tree nuts, and soybeans, make up the majority of the cupin superfamily. The prolamin superfamily consists of several important types of allergens in fruits, tree nuts, vegetables, legumes, and, cereals, that is the nonspecific lipid transfer proteins, the 2S albumin seed storage proteins, and the cereal a-amylase and protease inhibitors (Breiteneder et al., 2004)

Mainly, there are three protein allergens of cashew are identified by the WHO/IUIS i.e. Ana o 1, Ana o 2, and Ana o 3. These three allergens belong to the two superfamilies of the proteins - cupin (Ana o 1 and Ana o 2) and prolamin (Ana o 3). Specifically, Ana o 3 is a 2S albumin, Ana o 2 is an 11S globulin similar to that found in legumes, and Ana o 1 is a vicilin (Table 1). Ana o1 cashew allergen is heat and proteolysis resistant. De Leon et al in 2003, shows the cross-reactivity between the Ara h 1 and Ana o 1 allergens, and cashew shows there cross-reactivity with the pistachio. Willison et al in 2008, identified that cashew is the primary sensitizing agent in the pistachio/cashew allergic person.

Ana o 2 cashew allergen is legumin having a molecular weight of approx. 55 kDa and 457 aa long. Robotham et al in 2009, show the similarity and identity between 74% and 42% with 11S globulins, i.e. Ara h 3 of peanut, Gly m 6 of soybean, Cor a 9 of hazelnut, Ses i 7 of sesame, and Jug r 4 of walnut.

Ana o 3 is the 2S albumins having a molecular size below 14 kDa and this allergy affects the person due to their high immunoreactivity with sera. Ahn et al in 2009 identified the similarity and identity between pistachio allergen with 2S albumins and 11S globulin families.

Superfamily	Family	Biological function	Cashew
Prolamin	2S albumin	High stability to thermal and enzymatic treatment	Ana o 3
Cupin	Vicilins	Intermediate stability to thermal and enzymatic treatment	Ana o 1
	Legumins		Ana o 2

Table 1: Overview of family and biological function of allergenic cashew proteins

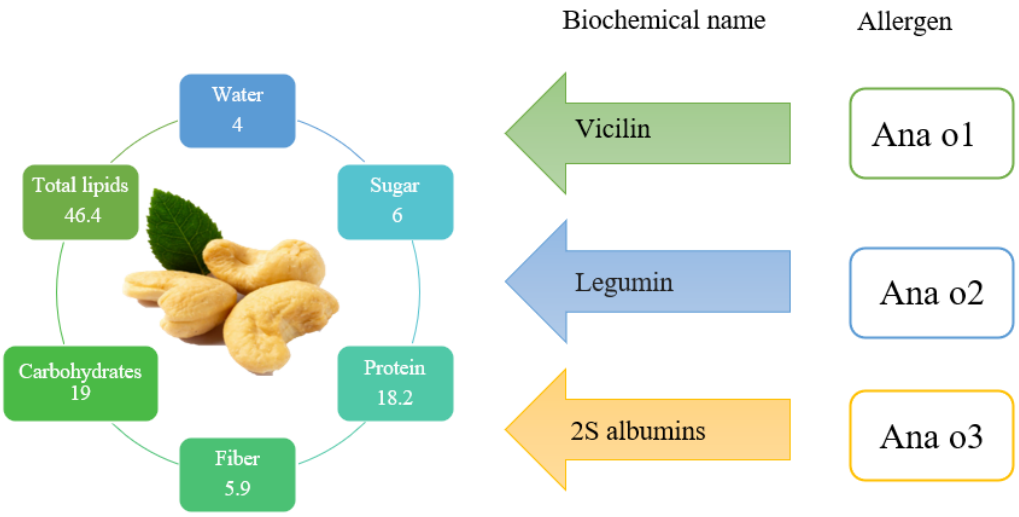


Figure 3. (**Left**) relevant information on the nutritional composition (g/100 g) relating to the *Anacardium occidentale* species. (**Right**) list of allergenic proteins identified and registered to date on the WHO/IUIS online allergen platform.

2.8 Prevalence of cashew allergy

Cashew nut allergy vary from one country to another, but peak in the USA *(*McWilliam et al., 2015). Around 5% of young children and 3-4% of people in developed nations suffer from food allergies. According to previously reported cases analysis, 90% of food-allergic reactions in children are because of items such as milk, egg, peanut, soy, and wheat and 85% of the food-allergic reaction are due to consumption of shellfish, tree nuts, fish, and peanuts in youngsters (Lee et al ., 2006; Hefle et al., 1996). Tree nut allergy is reportedly seen in up to 0.05% to 4.9% of individuals worldwide, according to research from a systematic review. (McWilliam et al., 2015). In the US cashew nut allergy come under the second highest incident rate among the various tree nut allergy (Fleischer et al., 2005). Fleischer et al in 2005 show that cashew nut allergies have a 30% prevalence in double-blind placebo-controlled food challenge (DBPCFC). Which is preferably done for tree nut allergy analysis. cashew allergy comes in the sixth position for single food allergy which can cause anaphylaxis in children and adults in European countries (Worm et al., 2014; Grabenhenrich et al., 2016). Davoren et al., 2005 shows that cashew allergy can be dangerous and serious and can cause the anaphylaxis in two-third of patients.

Even in modest doses, cashew allergies can cause dangerous reactions. In conclusion, numerous investigations have demonstrated that cashew allergies are associated with a high risk of anaphylaxis.

2.9 Epidemiology

Due to various factors, it is challenging for food allergies to become prevalent and dominant. Mainly IgE-mediated reactions are due to the 170 foods. Over time, variations had been seen in the frequency and various effects of food allergies. Indeed, several reports have shown to be effective in increasing the consequences during the last 10 to 20 years. Generally, various reports show the consequences of food allergy, but the most common tree nut allergy is cashew allergy (Sicherer et al., 2001). Davoren et al in 2005 revealed that children and infants have cashew allergies to a greater extent than adults. Around 0.08% of infants (age less than 4) in the UK were suffering from cashew allergy. Hasegawa et al in 2009 identified female adults to have a cashew nut allergy, there are some consequences of cashew allergy in Asia. Out of 100 people UK 41 % of individuals are allergenic which includes cashew allergy. Cashew allergy mainly occurs when they are highly intake of Asian infants. urushiol dermatitis, systemic

dermatitis and serious allergic interaction dermatitis are caused by the Anacardiacea family. Despite the perception that cashew nut sensitization and clinical allergy are rising, methodologically relevant studies proving this has not yet been carried out.

2.10 Clinical Features

Cashew allergy is increasing day by day and can have fatal anaphylactic reactions to the patient (Tufail et al., 2019). There is some paper published on cashew allergy, showing that the person having allergy have skin diseases, gastro-intestine, and respiratory problems (Valk et al., 2014). The 1/3rd part of the saturated fatty acid in the cashew is stearic acid, which is neutral in blood lipid, due to which cashew effects the human body as similar to other nuts. Thus, the data on blood lipids and cashew have been limited (Mah et al., 2017). Mohan et al in 2018, discussed that cashew nut consumption's decreases the levels of low-density lipoproteins, triglyceride, and cholesterol in the blood and increases the levels of high-density lipoproteins in the serum. Mah et al in 2017, reported that consumption of cashew nuts is associated with the reducing the risk of the type II diabetes and it also effect the cholesterol level in human body. Recent studies have also shown an interest in the antioxidant properties of the anacardic acids found in cashew nuts, with results pointing to possible therapeutic use for certain diseases. These have a potential neuroprotective effect against degenerative changes in Parkinson's disease or they are linked to the protection of behavioural alterations and oxidative stress generated by rotenone in a rat model of the disease (Medeiros-Linard et al., 2018).

Davis et al in 2007, identified that adding cashew to the diet increases the antioxidant capacity and results in metabolic syndrome. Furthermore, Estruch et al in 2013, figured out that the consumption of cashew have benefit in front of a low-fat diet otherwise cashew nut causes heart disease due to fatty acids.

2.11 Methods for Detection of Cashew Allergenicity

2.11.1 Immunological/ Molecular Biology Analytical Methods for Cashew Detection

For the identification and detection of cashew allergens from the foodstuffs there are several methods which include the immunological and molecular biology analytical methods like

ELISA. Mendes et al in 2019 demonstrated the sensitivity of 0.2–1 mg/kg or 1–2 mg/kg of cashew in food matrices with ELISA and Lateral Flow Devices (LFD) kits respectively. Along with it, also identify some of the cross-reactivity between the cashew and other tree nuts including pistachio, peanut, Brazil nut, walnut and hazelnut.

Similarly, Wei et al in 2002 experimentally detected the cashew traces in different foodstuffs like milk chocolate, wheat flour, rice cereals, rolled oats, chocolate-filled cookies and raisin bran cereal through the sandwich ELISA (commercial kits). In addition to it, Ben et al in 2005 also identified the cashew in dark chocolate and milk by the multitarget indirect ELISA method at 1 mg/kg level. Furthermore, Zhao et al in 2019 developed two different sensitive sandwich ELISA for the albumin 2S (Ano o 3) which acts as the stable protein marker for the identification of the residues of the cashews in prepackaged foodstuffs mainly in processed cookies and chocolate, but for the Ana o2, sensitive, sandwich ELISA method gives the less cross-reactivity of cashew Ano o2 protein with pecan, hazelnuts, almonds, peanuts and pistachio.

Real-time PCR assay is also one of the detection methods for cashew in the different processed and raw food matrices such as cookies, spreads, chocolate, pesto and ice creams (López et al., 2015). Mainly, the protein-based method and the advanced DNA approaches were focused on the identification of the single copy of cashew genes i.e., Ana o 1, Ana o 2, Ana o 3, which doesn't show the cross-reactivity with other animal and plants species (Prado et al., 2016). To detect the cashew proteins in the various foodstuffs there are multi- target method with high sensitivity and specificity.

2.11.2 Mass Spectrometry-Based Analytical Method for Cashew Analysis

For the analysis of the allergy, mass spectrometry is one of the methods to be used. Bignardi and coworkers in 2010, developed the LC-MS method, which is one of the methods of mass spectrometry which have been used for the quantification and the detection of cashew nuts in the food (Luparelli et al., 2022). Bignardi et al., 2010 detect the Ana o 2 cashews in many food products like biscuits, cakes, chocolate and flour. Korte et al in 2016 proposed some of the new protein-peptide markers for the cashew by the LC-HRMS multi-method (Korte et al., 2016). Moreover, three cashew protein peptide markers were depicted from the protein sample of ice cream.

Korte et al in 2019 identified the three-cashew protein peptide from the food matrices for the detection of the allergen marker. It has been demonstrated that the potential interference with

the particular food matrices, degradation and rate of tryptic peptide synthesis might affect the allergen quantification and should be taken into consideration when choosing the markers, such as 20-83% of allergen properties lost after baking biscuits and breads as the heat treatment has the significant impact on the allergen detectability. (Korte et al., 2019).

Using the UHPLC-MS/MS method, cashew allergens were discovered in complex food like ice cream and chocolate as well as in processed foodstuff including sauces and cookies (Planque et al., 2017). On the other side, Gu et al in 2018 with the help of a quadrupole-Orbitrap instrument, which is also mass spectrometry predicted that using the marker of Ana o 2 cashew protein peptide i.e, ADIYTPEVGR gives the 0.7 µg/g and 2.3 µg/g value of LOD and LOQ.

New et al in 2018 identified the 12 allergens which also includes the cashew allergens with the help of the LC-MS method and also predicted the parameters for the validation, screening, quantification and identification of the 12 allergens by selecting the specific peptide for every allergen at the 10 µg/g detection limit in bread and cookies which contain cashew and peanuts (New et al., 2018).

2.11.3 Computational Tool for the Study of Allergens

Conventional laboratory methods are used for the detection and study of allergens, but complementary to this there are many useful bioinformatics tools for the prediction of allergenicity (Nedyalkova et al., 2023*)*. Computational tools were used for sequenced and structural-based analysis of allergens, B- cell and T-cell epitope prediction, genome analysis, cross-reactivity assessment and many more (Brusic et al., 2003*)*. To examine the databases, these computational tools are used for the characterization and classification of allergens and genes - comparison and sequence analysis tools are used, similarly for the structural and functional analysis of proteins and genes to predict the putative allergens-prediction tools are preferred (Lingyi et al., 2023*)*. Some of the bioinformatics tools are described in Table 2. Moreover, the prediction of highly accurate data can enable in silico validation of putative allergens more rapid and cost-effective. For the B-cell and T-cell epitope prediction *(*Armina et al., 2019), to check the cross-reactivity of allergens and allergenicity prediction, bioinformatics tools are used (María et al., 2023). In addition, the secondary and tertiary structure can also be predicted through these tools which enables to study of the structural properties of the putative allergens (Srishti et al., 2022).

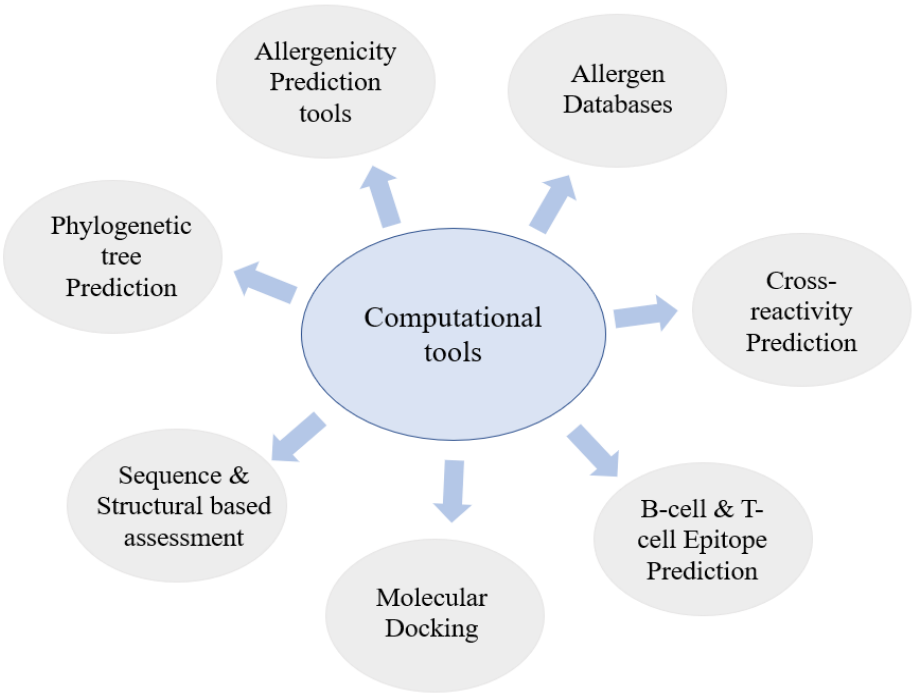


Figure 4: Different Computational Tools Used for Allergen Prediction

Name	Link (Website)	Description	References
Allergen nomenclature	http://www.allergen.org	Official site for the systematic allergen nomenclature provided by the World Health Organization and International Union of Immunological Societies (WHO/IUIS)	Pomés et al., 2018
AllergenOnline	http://www.allergenonline.org	Provides sequence database of allergens to identify proteins and assess the potential risk of allergenic cross-reactivity. This database offers 2233 peer-reviewed sequences from 912 taxonomic protein groups (February 2021)	Goodman et al., 2016
Allergome	http://www.allergome.org	A website with detailed information on Allergenic Molecules (Allergens) causing an IgE-mediated (allergic, atopic) disease (anaphylaxis, asthma, atopic dermatitis, conjunctivitis, rhinitis, urticaria)	Mari et al., 2006
Database of Allergen Families-AllFam	http://www.meduniwien.ac.at/allfam/	Comprises a resource for classifying allergens into protein families as well as biochemical properties and allergology significance	Radauer et al., 2008
Immune Epitope Database and analysis resource (IEDB)	https://www.iedb.org	Provides experimental data on antibody and T-cell epitopes to identify allergens and to assist in the prediction and analysis of allergenicity	Vita et al., 2019
Structural Database of Allergenic Proteins (SDAP)	https://fermi.utmb.edu	Tool for testing the FAO/WHO allergenicity rules in new proteins and investigating cross reactivity, also offering information about protein sequence and structure	Schein et al., 2022
Allermatch	http://www.allermatch.org	webtool where you can compare the amino acid sequence of a protein of interest with sequences of allergenic proteins	Fiers et al., 2004
AllerTOP	https://www.ddg-pharmfac.net/AllerTOP/	The Auto Cross Covariance (ACC) transformation approach is used to define allergenicity in AllerTOP	Dimitrov et al., 2014
AllergenFP	http://ddg-	Tool accounts Tanimoto coefficients (T values), which range from 0 to 1 to	Dimitrov I et al., 2014

	pharmfac.net/AllergenFP/method.html	predict allergenicity	
AlgPred	http://crdd.osdd.net/raghava/algpred/	Tool use SVM support vector machine algorithm to predict allergenicity	Saha et al., 2006
Web Gene Ontology Annotation Plot server. (WEGO)	www.wego.genomics.cn	Tool for visualization and plotting of gene ontology data, accession numbers were submitted to WEGO to plot a graph	Ye et al., 2018
I-TASSER 3D	https://zhanglab.ccmb.med.umich.edu/I-TASSER/	Tool for the structure prediction which is used to analyse the sequences of allergen proteins	Yang et al., 2015
ClusPro 2.0	www.cluspro.bu.edu/home.php	web service for protein-protein and protein-peptide docking	Desta et al., 2020

Table 2: Some Bioinformatic software tools most used for allergen analysis

3.1. Sequenced-based assessment

3.1.1 Sequence retrieval

Cashew protein sequences were retrieved from the UniProt database (https://www.uniprot.org/), which served as the query sequence. Similarly, allergens related to food were identified and retrieved from the World Health Organization and International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-committee server (http://www.allergen.org/) (Singh et al., 2023).

3.1.2 Sequence based allergenicity prediction

Retrieved query sequences of Cashew protein were subjected to alignment against retrieved food allergen sequences from the allergen.org database using pBLAST (packaged in Bioedit software, v 7.2.5) by using the default algorithm parameter. Results of pBLAST were obtained in the form of E-value, percent identity and query coverage. Sequences having query coverage and percent identity >50% were subjected against three different allergen databases namely Allermatch (http://www.allermatch.org) , SDAP (https://fermi.utmb.edu/) and AllergenOnline (http://www.allergenonline.org) by using a method of Full FASTA 35 and 80 amino acid sliding window of Full FASTA 35. The alignment was executed by using the default parameters of the tool. AllergenOnline tool uses BLOSUM 50 as a scoring matrix to perform alignment, results were generated and sequences having percent identity>50% and E-value of 1 in Full FASTA mode and percent identity >35% in sliding window of 80 amino acids were selected and used for further analysis. The final result was generated using the BLAST algorithm, shortlisted sequences of a query of cashew protein with food allergen sequences to select consensus sequences (vashisht et al., 2023).

3.1.3 Physicochemical properties based allergenicity prediction

For physiochemical properties based on allergenicity prediction for cashew protein we used the approach given by vashisht et al in 2016 to assess the allergenicity of shiitake mushrooms. Consensus sequences were retrieved from Uniprot (https://www.uniprot.org/) which were

showing the cross- reactivity with the food allergens and these sequences were further used to predict the allergenicity using the for 3 different servers: AllerTOP (https://www.ddg-pharmfac.net/AllerTOP/), AllergenFP (http://ddg-pharmfac.net/AllergenFP/method.html), and AlgPred (http://crdd.osdd.net/raghava/algpred/) with default parameters setting. The AlgPred tool uses a SVM support vector machine algorithm to predict allergenicity, tool can assess amino acid composition, motif and IgE epitopes for allergenic sequences. To access the allergenicity of protein sequence tool considered different properties of protein sequence like charge, side chain, hydrophobicity and hydrophilicity, and amino acid composition. The Auto Cross Covariance (ACC) transformation approach is used to define allergenicity in AllerTOP and AllergenFP. AllergenFP accounts for Tanimoto coefficients (T values), which range from 0 to 1 to predict allergenicity, a higher T value linked with either an allergic or non-allergenic query sequence was deemed more authentic. The terms "probable allergen or probable non-allergen" were used to describe the results in all 3 tools, and details about the closest protein were also provided. On the bases of the consensus method of all these three servers, the potential allergenicity of the shortlisted proteins was finally predicted.

3.1.4 Sequence alignment using the MUSCLE tool

MUSCLE stands for multiple sequence comparison by log expectation. MUSCLE tool is used to perform multiple sequence alignment of protein sequences, it can be used to access evolutionary relationships and degree of conservancy between protein sequences. For estimating the degree of conservancy entire length of the query sequence was taken into account, only identical amino acid residues of the query sequence with functional similarity were taken to calculate the score of conservancies from CLUSTAL W output format.

3.2. Structure based allergenicity assessment

3.2.1 Secondary structure prediction

Using PHD and PSIPRED fold servers, secondary structures of a subset of allergen proteins were predicted, and a comparison was conducted (Vashisht et al., 2023).

3.2.2 Three-dimensional structure-based allergen prediction

The I-TASSER 3D structure prediction program is used to analyse the sequences of allergen proteins (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). The server was utilized with the

default parameters, and the results were examined. Models with acceptable values for several parameters including C-score, B factor, RMSD values, and, T.M score were chosen, and SAVES (http://servicesn.mbi.ucla.edu/SAVES/) was used to assess stereo-chemical quality and reliability. ModLoop (https://modbase.compbio.ucsf.edu/modloop/) was used for loop refining to improve and redesign the poorly defined loop portions. The ProSA server was used to examine the Z-score of the acquired structure (https://prosa.services.came.sbg.ac.at/prosa.php). With the help of the UCSF CHIMERA program (https://www.cgl.ucsf.edu/chimera/), prediction accuracy analysing was done which finalized food protein models were superimposed with templates, followed by superimposition with that of potent food allergens. Lower RMSD values suggested a greater structural resemblance between the known allergen and the computationally created model of the Cashew proteins.

3.2.3 Classification of the allergen sequence along with their gene ontology

Screened query sequences of cashew were submitted to Interpro (www.ebi.ac.uk/interpro) and conserved domain database (www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) to analyse a family of them using Hidden Markonikof Model Method. Further gene ontology accession number was carried out using the Interpro database to determine their molecular, biological and cellular function. For visualization and plotting of gene ontology data, accession numbers were submitted to the Web Gene Ontology Annotation Plot server. (WEGO, www.wego.genomics.cn)

3.2.4 B-cell epitope prediction

The B-cell epitope is the antigen portion binding to an antibody. For the prediction of B-cell epitopes The Immune Epitope database was used (https://www.iedb.org/). This database catalogue's experimental data on T-cell epitopes, antibodies, and autoimmune disorders from human and other animal species. This database contains various tools for the prediction and analysis of epitopes. To sequentially characterise antigenic epitopes utilising amino acid input sequences, a variety of techniques are used in linear B-cell epitope prediction. With start and end regions, it will produce potential epitopes. For further computational analysis, the final epitopes can be chosen from the predicted epitopes (Mishra et al.,2016).

3.2.5 Structure Modelling

An automated system called Swiss-model (http://swissmodel.expasy.org/) uses homology modelling techniques to model the protein 3D structure from its amino acid sequence. Swiss Model is a web server run by the University of Basel's Swiss Institute of Bioinformatics Biozentrum (Waterhouse ,2018), with the server's user-friendly web interface, even non-specialists can create 3D models of their chosen proteins using just a standard web browser. To begin a fresh modelling project, click the "Start Modelling" button. You can either paste the target protein's amino acid sequence or submit it in FASTA or plain text format. You can also enter the target sequence's UniProtKB AC in the input field. The "Build Model" option launches an automatic pipeline that searches for templates. The automated mode chooses templates to maximise the expected model quality. When the search for templates is complete, the output page comprises a main table listing the available templates and ranking them based on the expected quality of the resulting models. For further analysis, the model with the highest quality index was chosen. Download the model in PDB format. The ProCheck tool was used to evaluate the structure's quality (Singh et al., 2022).

3.2.6 Molecular Docking

The process of modelling the structure produced by two molecules interacting is known as docking. Protein docking, one of the most studied topics in computational and structural biology, has several applications, including drug development. For molecular docking, ClusPro 2.0 (www.cluspro.bu.edu/home.php) is used, which is a web service for protein-protein and protein-peptide docking. In Boston University's Vajda lab, this server was built and is still maintained. The server requires two Protein Data Bank (PDB) files in protein-protein mode or a PDB file for the protein and a sequence for the ligand (protein-peptide mode) as input. Finally, an antibody structure which had been retrieved from a protein data bank was uploaded as a receptor and a potential allergen structure that had been chosen was uploaded to the server as a ligand. Its output is made up of approximately 10 models of the final structure that is formed by the interaction of the two components. Results are normally generated by the server in around four hours. The server also gives users the ability to fine-tune the results using any extra information about the interaction process, such as the small-angle X-ray scattering (SAXS) profile or distance restraints, via the "Advanced Options" section. To examine binding interactions, the position with the highest docking score was chosen. The conformation with the best score was chosen, and PyMol analysis was performed on it. PyMol was used to analyse

interactions and visualise docked complexes in three dimensions. PyMOL was loaded with a protein structure file, typically one from the PDB file, that was retrieved from ClusPro 2.0, and the structure was then visualised (Bonvin et al., 2006).

3.3 Expression analysis of identified allergens by using putative available transcriptomics data of *Anacardium occidentale*

After obtaining the high-quality pre-processed data, the next step is to align the reads against the reference data. Aligning the reads against the reference data helps to know which transcripts are expressed by mapping the reads to the reference data (genomic mapping or transcriptomic mapping) (Karnaneedi et al., 2020).

Transcriptomic data of *Anacardium occidentale* was downloaded from the NCBI SRA toolkit (Project ID- SRR20532291, Accession: SRX16556127). The gene sequences of the 11 allergen proteins were downloaded using NCBI and UniProt. These sequences were aligned against the transcriptomic data of *Anacardium occidentale* using BLAST alignment to find the gene expression between transcriptomic data and the allergen sequences. It returned the output in the form of a hit table with the query coverage, E-value and percentage identity. Mainly, the reliability of the alignment is checked through E-value (Karnaneedi et al., 2020).

4.1 Cross-reactivity with food allergens

Cashew sequences were obtained from the UniProt database 114 which served as the query sequence. Similarly to this, 476 food allergens were retrieved from the server of the Allergen Nomenclature Sub-committee of the World Health Organization and International Union of Immunological Societies (WHO/IUIS), which serves as the database. These sequences are further examined for sequence alignment using the FASTA and BLAST tools. First, following BLAST alignment, 54 sequences were shortlisted based on cutoff values of >50% (percent identity), 1.0E-07 (E-value) for the FASTA alignment, these sequences were evaluated using two allergen databases (Allermatch and Allergen online) having criteria as % identity > 50% for Full FASTA alignment and 35% for the 80 amino acid sliding window FASTA alignment. The Pblast algorithm was then applied to 30 hits that showed results on these two platforms. A consensus method was used to narrow down the list of proteins. Thus, only 11 out of 30 proteins (NAD(P)H-quinone oxidoreductase subunit K, Photosystem II reaction center protein K, 30S ribosomal protein S11, ATP-dependent Clp protease, 30S ribosomal protein S7, and Pathogenesis-related protein 10), showed consensus in these two allergen databases and plausible BLAST alignment results, thereby predicted to have potential cross-reactivity with potent food allergens.

4.2 Physicochemical properties based allergenicity prediction

Eleven shortlisted proteins (NAD(P)H-quinone oxidoreductase subunit K, Photosystem II reaction centre protein K, 30S ribosomal protein S11, ATP-dependent Clp protease, 30S ribosomal protein S7, and Pathogenesis-related protein 10) were evaluated for physicochemical allergenic potential- based assessment using three *In silico* tools: AllergenFP, AllerTOP and AlgPred. Using SVM (Support Vector Machine) based amino acid/dipeptide composition method and Antigen Representing Peptides (ARP) BLAST, AlgPred predicted, 6 out of 11 proteins to be allergens. Similarly, proteins, respectively, were predicted to be allergens using AllerTOP and AllergenFP. Along with the Tanimoto coefficient, which ranged from 0.79 to 0.85, information about proteins with a high degree of sequence similarity to the query was also given (Table 3). So, using a consensus approach, 8 proteins (including NAD(P)H-quinone oxidoreductase subunit K, 30S ribosomal protein S11, and Pathogenesis-related protein 10)

were chosen and predicted to have allergenic characteristics based on their physicochemical properties. It was predicted that HSP 70 and trans aldolase have a non-allergenic nature.

Protein Sequence (Accession number)	**AlgPred**				**AllerTOP**		**AllergenFP**			**Consensus Result**
	SVM amino acid composition	**SVM dipeptide composition**	**ARP BLAST**	**Hybrid method**	**Result**	**Nearest allergen (Accession number)**	**Result**	**Tanimoto index**	**Nearest allergen (Accession number)**	
NAD(P)H-quinone oxidoreductase subunit K **(A0A1Z1G942)**	Non allergen	Non allergen	Non Allergen	Non Allergen	**Probable allergen**	dipeptidylpeptidase IV preproprotein [Vespula vulgaris]. (ACA00159.1)	**Probable allergen**	0.79	VIT1_CHICK Vitellogenin-1 (P87498)	**Allergen**
Photosystem II reaction center protein K **(A0A1Z1G912)**	Non allergen	Non allergen	Non allergen	Non allergen	Probable Non-allergen	CYB_SOLTU Cytochrome b (P29757)	Probable Non-allergen	0.84	PUT1_ORYSJ Polyamine transporter PUT1 (Q6Z8D0)	Non- allergen
30S ribosomal protein S11 **(A0A1Z1G958)**	Non allergen	Non allergen	Non allergen	Non allergen	**Probable allergen**	60S ribosomal protein L3 (Allergen Asp f 23). (Q8NKF4)	Probable Non-allergen	0.83	CARF_HUMAN CDKN2A-interacting protein (Q9NXV6)	**Allergen**

ATP-dependent Clp protease **(A0A1Z1G961)**	Non allergen	Non allergen	Non allergen	Non allergen	Probable Non-allergen	CA2D4_HUMAN Voltage-dependent calcium channel subunit alpha-2/delta-4 (Q7Z3S7)	Probable Non-allergen	0.8	C5AR2_HUMAN C5a anaphylatoxin chemotactic receptor 2 (Q9P296)	Non- allergen
30S ribosomal protein S7 **(A0A1Z1G996)**	Non allergen	Non allergen	Non allergen	Non allergen	Probable Non-allergen	CCD22_HUMAN Coiled-coil domain-containing protein 22 (O60826)	Probable Non-allergen	0.82	BPTF_HUMAN Nucleosome-remodeling factor subunit BPTF (Q12830)	Non- allergen
Pathogenesis-related protein 10 **(A0A7D9N3X7)**	**Potential allergen**	**Potential allergen**	**Allergen**	**Allergen**	**Probable allergen**	MALDO Major allergen Mal d 1 (Q43549)	**Probable allergen**	0.84	Q6QHU2_PRUAV Major cherry allergen Pru av 1.0202 (Q6QHU2)	**Allergen**
Pathogenesis-related protein 10 **(A0A7D9N3Y2)**	**Potential allergen**	**Potential allergen**	**Allergen**	**Allergen**	**Probable allergen**	Q43549_MALDO Major allergen Mal d 1 (Q43549)	**Probable allergen**	0.85	Q43551_MALDO Major allergen Mal d1 (Q43551)	**Allergen**
Pathogenesis-related protein 10	**Potential allergen**	**Potential allergen**	**Allergen**	**Allergen**	Probable Non-allergen	BTBD1_HUMAN BTB/POZ domain-containing protein 1	**Probable allergen**	0.82	pollen allergen Que a 1 isoform [Quercus alba] (ABZ81046)	**Allergen**

(A0A7D9N3Y7)						(Q9H0C5)				
Pathogenesis-related protein 10 **(A0A7D9N3Z9)**	**Potential allergen**	**Potential allergen**	**Allergen**	**Allergen**	**Probable allergen**	Cas s 1 pollen allergen [Castanea sativa]. (ACJ23863)	**Probable allergen**	0.84	Q43551_MALDO Major allergen Mal d1 (Q43551)	**Allergen**
Pathogenesis-related protein 10 **(A0A7D9N402)**	**Potential allergen**	**Potential allergen**	**Allergen**	**Allergen**	**Probable allergen**	major allergen Cor a 1 [Corylus avellana]. (CAA96549)	**Probable allergen**	0.83	Ribonuclease-like PR-10c (Mal d 1.0109) (Q941P6)	**Allergen**
Pathogenesis-related protein 10 **(A0A7D9N413)**	**Potential allergen**	**Potential allergen**	**Allergen**	**Allergen**	Probable Non-allergen	BTBD1_HUMAN BTB/POZ domain-containing protein 1 (Q9H0C5)	**Probable allergen**	0.82	pollen allergen Que a 1 isoform [Quercus alba] (ABZ81046)	**Allergen**

Table 3: Combined physico-chemical property-based allergenicity result of 11 cashew proteins derived using AlgPred, AllerTOP and AllergenFP servers

4.3 Sequence alignment using MUSCLE tool

Using the MUSCLE tool for sequence alignment, the amino acid sequences of all 8 allergen proteins were aligned with the functionally related food allergens. All of the proteins had a degree of conservancy ranging from 30.4%-66.8% (Table 4). The sequence alignment image between the allergen Pru av 1.0201 (Q6QHU3) and the cashew protein Pathogenesis-Related Protein 10 (A0A7D9N3Y2) is shown in Figure 5. The existence of identical amino acids or their substitution with functionally conserved amino acids in the query protein suggested the sequence conservancy between two sets of proteins using the formula given below, therefore, suggesting cross-reactivity and potential allergenicity between protein sequence and their respective food allergen.

Degree of conservancy = Identical amino acids (*) + substitute amino acids (:)/total no. of amino acid residues throughout the entire length X 100

Protein Sequence (Accession number)	**Length of amino acid sequence**	**Food allergen (Accession number)**	**Length of amino acid sequence**	**Degree of sequence conservancy**
NAD(P)H-quinone oxidoreductase subunit K **(A0A1Z1G942)**	227	Pru p 1 (Q2I6V8)	**160**	**30.8**
30S ribosomal protein S11 **(A0A1Z1G958)**	138	ONCKE (D5MU14)	**193**	**30.4**
Pathogenesis-related protein 10 **(A0A7D9N3X7)**	154	Pru p 1.0201 (AJE61290.1)	**160**	**66.8**
Pathogenesis-related protein 10 **(A0A7D9N3Y2)**	154	Pru av 1.0201 (Q6QHU3)	**160**	**66.25**
Pathogenesis-related protein 10 **(A0A7D9N3Y7)**	154	Pru ar 1 (O50001)	**160**	**65**
Pathogenesis-related protein 10 **(A0A7D9N3Z9)**	154	Pru av 1.0202 (Q6QHU2)	**160**	**62.5**
Pathogenesis-related protein 10 **(A0A7D9N402)**	159	Pru p 1.0301 (AJE61291.1)	**160**	**-**

Pathogenesis-related protein 10 **(A0A7D9N413)**	155	Mal d 1 (Q43550)	**160**	**62.5**

Table 4: Sequence alignment-based conservancy results for 8 selected cashew proteins with respect to corresponding food allergens

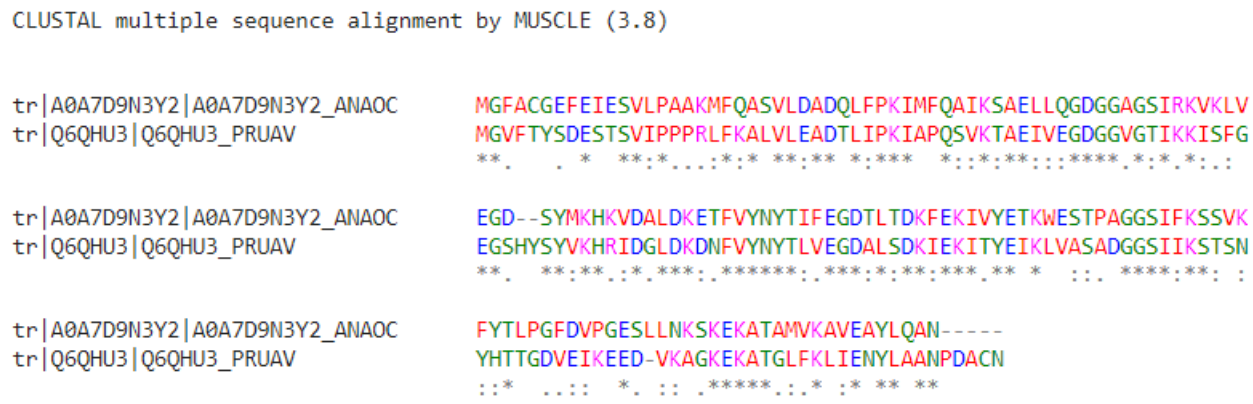


Figure 5: Sequence alignment of Pathogenesis-related protein 10 **(A0A7D9N3Y2)** and food allergens Pru av 1.0201 (Q6QHU3)

4.4 Secondary structure prediction

Using PHD, PSIPRED server, the secondary structure of the cashew protein sequence and associated food allergies were determined. As shown in supplementary table 4, a comparative assessment of these proteins was established. Through the use of the PSIPRED and PHD tools, the secondary structure between the allergen Pru av 1.0201 (Q6QHU3) and the cashew protein Pathogenesis-related protein 10 (A0A7D9N3Y2) is shown in Figs. 6 and 7 respectively. Comparing cashew proteins to food allergens revealed similarities in secondary structure and amino acid composition, pointing to potential cross-reactivity between the two.

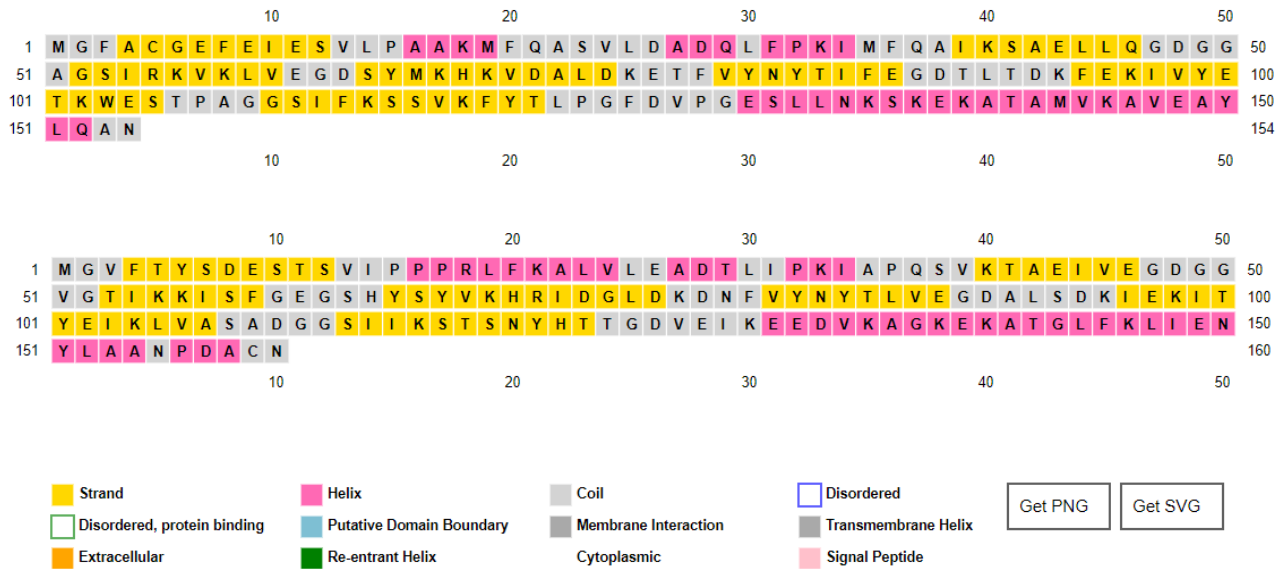


Figure 6: Secondary structure based comparative analysis of Pathogenesis-related protein 10 **(A0A7D9N3Y2)** and food allergens Pru av 1.0201 (Q6QHU3) using PSIPRED server

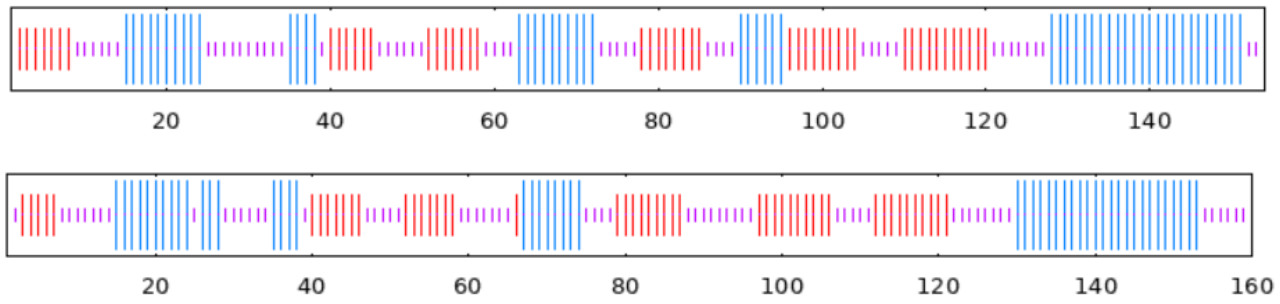


Figure 7: Secondary structure based comparative analysis of Pathogenesis-related protein 10 **(A0A7D9N3Y2)** and food allergens Pru av 1.0201 (Q6QHU3) using PHD serve

4.5 Three-dimensional structure prediction

Using specific templates produced from Pblast analysis, the I-TASSER server was used to create three-dimensional models for each protein. Models with appropriate B factor, TM value, C score, and RMSD values were chosen and put through a quality assessment utilizing the SAVES 6.0 server. B factor is known as the normalization of the predicted Z-score. The B factor value must be zero or predominantly negative (≤0) for a stable structure. The confidence

score for the efficiency of the predicted model is considered to be C- score. The structure would be considered to be more signifying reliable when its C- score range is between -5 to +2. The topological similarity between the template and the anticipated model is indicated by the TM and RMSD values. Using the ModLoop server, loop refinement was carried out for residues with inadequate modelling. Figure 8, represents Three dimensional structure modelling for cashew protein- Pathogenesis-related protein 10 **(A0A7D9N3Y2)** using I-TASSER server, whereas other cashew proteins are also depicted

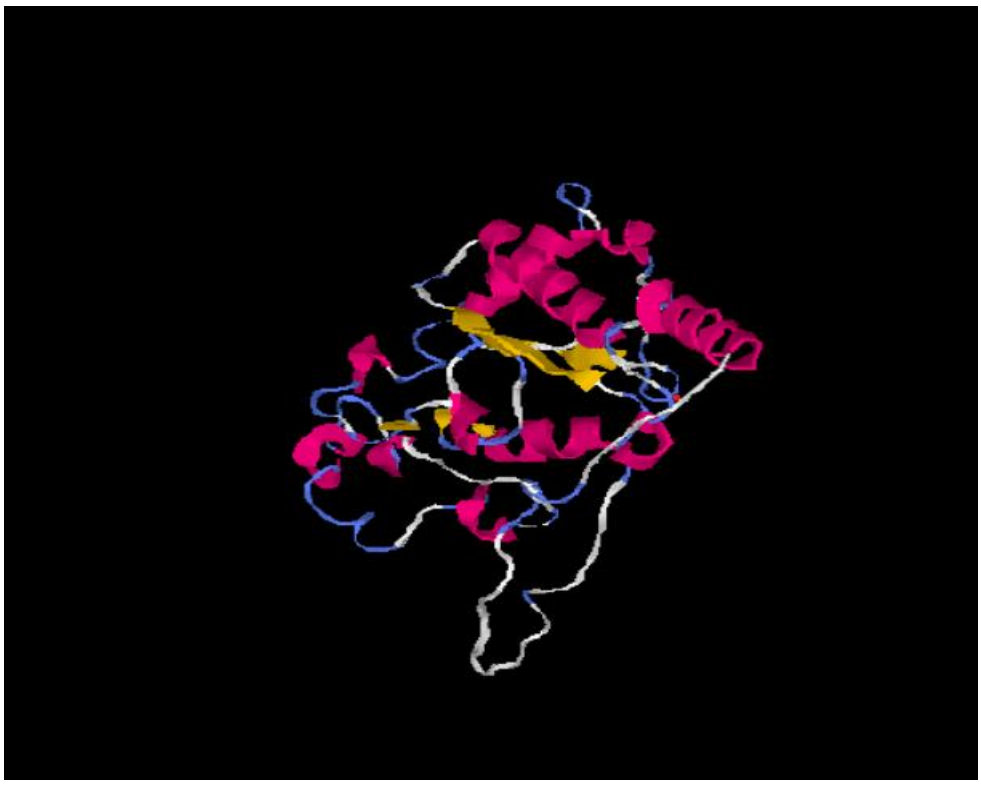


Figure 8 : Three dimensional structure modelling of Pathogenesis-related protein 10 **(A0A7D9N413)** and food allergens Mal d 1 (Q43550) using PHD server

4.6 Classification of the allergen sequence along with their gene ontology

Protein family and gene ontology analysis was performed on the selected 8 cashews allergen sequences. For protein family analysis, the Pfam database and the Conserved Domain Database were employed. Results are displayed in Tables 5 and 6, respectively. Allergen sequences were put via HMMER to determine the protein family hidden Markov model approach table 7. The gene ontology accession number was obtained from the InterPro website (table 8). Based on the GO accession number, the WEGO tool (Web Gene Ontology Annotation Plot) was used to visualize the gene ontology data (Graph 1).

Query	Family	Description	Entry type	clan	Envelope		Alignment		HMM		HMM length	Bit score	E – value	Predicted active site
					Srart	End	start	end	Start	End				
A0A1Z1G942	Oxidored_q6	NADH ubiquinone oxidoreductase,20 kd subunit	Family	n/a	42	153	45	153	5	129	129	79.6	1.9e-22	n/a
A0A1Z1G958	Ribosomal _S11	Ribosomal protein_S11	Family	CL0267	28	137	28	137	1	110	110	170.6	1.2e-50	n/a
A0A7D9N3X7	Bet_v_1	Pathogenesis related protein Bet v 1 family	Domain	CL0209	2	154	7	153	7	150	151	65.4	5.7e-18	n/a
A0A7D9N3Y2	Bet_v_1	Pathogenesis related protein Bet v 1 family	Domain	CL0209	2	154	7	153	7	150	151	63.5	2.2e-17	n/a
A0A7D9N3Y7	Bet_v_1	Pathogenesis related protein Bet v 1 family	Domain	CL0209	2	154	6	152	7	149	151	69	4.5e-19	n/a
A0A7D9N3Z9	Bet_v_1	Pathogenesis related protein Bet v 1 family	Domain	CL0209	2	154	7	153	7	150	151	66.6	2.4e-18	n/a
A0A7D9N402	_	_	_	_	_	_	_	_	_	_	_	_	_	_
A0A7D9N413	Bet_v_1	Pathogenesis related protein Bet v 1 family	Domain	CL0209	1	155	7	153	7	149	151	69.5	3.0e-19	n/a

Table 5: Pfam database result of 8 cashew protein allergens

Query	Hit type	PSSM-ID	From	To	E-Value	Bitscore	Accession	Short name	Incomplete	Superfamily
A0A1Z1G942	specific	214337	1	227	0	526.967	CHL00023	ndhK	-	cl17194
A0A1Z1G958	specific	176982	15	130	8.01658e-77	223.211	CHL00041	rps11	-	cl00332
A0A7D9N3X7	specific	176858	7	151	7.07392e-49	154.27	cd07816	Bet_v1-like	-	cl14643
A0A7D9N3Y2	specific	176858	7	151	7.89059e-49	153.884	cd07816	Bet_v1-like	-	cl14643
A0A7D9N3Y7	specific	176858	8	151	3.14196e-43	139.632	cd07816	Bet_v1-like	-	cl14643
A0A7D9N3Z9	specific	176858	7	151	2.01907e-48	153.114	cd07816	Bet_v1-like	-	cl14643
A0A7D9N402	specific	-	-	-	-	-	-	-	-	-
A0A7D9N413	specific	176858	9	152	2.3238e-43	140.017	cd07816	Bet_v1-like	-	cl14643

Table 6: Conserved Domain Database result of 8 cashew protein allergens

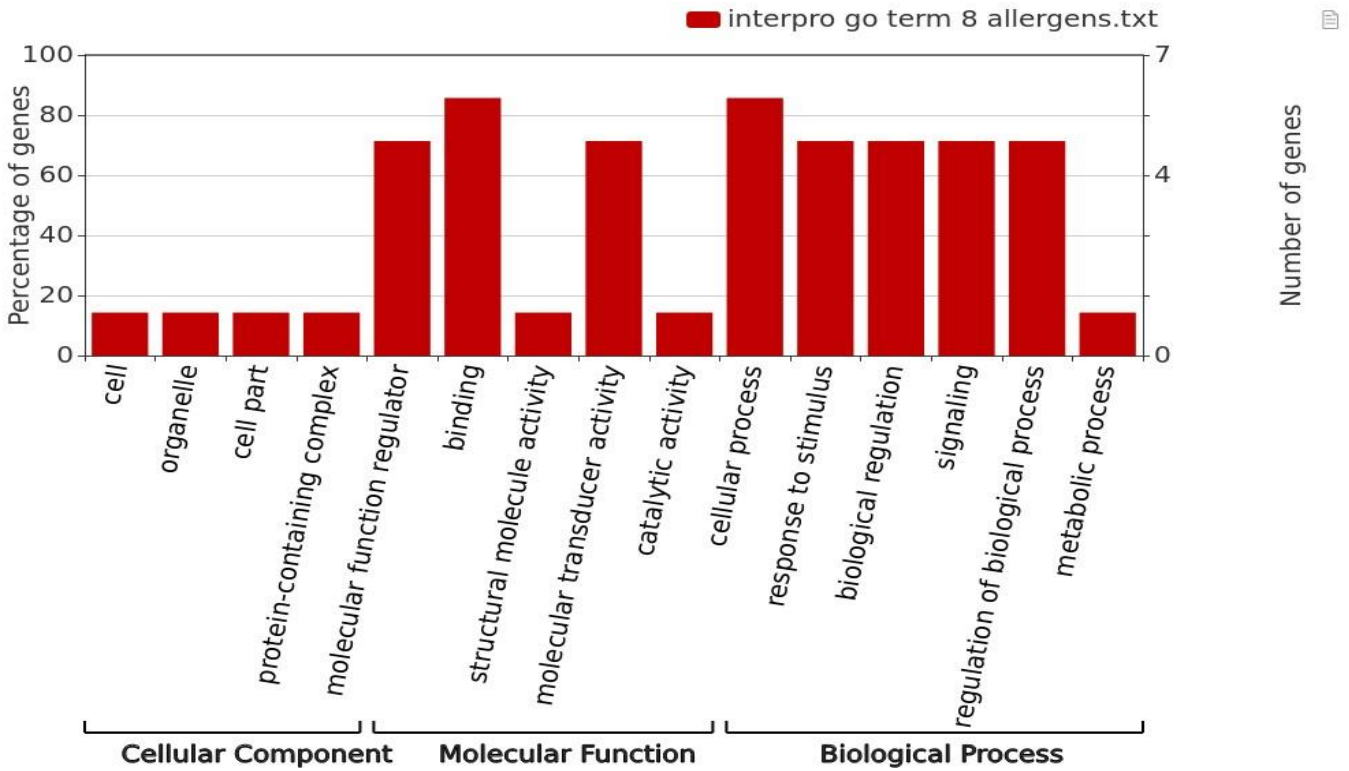
Query	Family		Clan	Description	Start	End	Domain E-value	
	Id	Accession					Ind.	Cond.
A0A1Z1G942	Oxidored_q6	PF01058.25	n/a	NADH ubiquinone oxidoreductase,20 kd subunit	42	153	1.9e-22	9.8e-27
A0A1Z1G958	Ribosomal _S11	PF00411.22	CL0267	Ribosomal protein_S11	28	137	1.2e-50	6.3e-55
A0A7D9N3X7	Bet_v_1	PF00407.22	CL0209	Pathogenesis related protein Bet v 1 family	2	154	5.7e-18	2.9e-22
A0A7D9N3Y2	Bet_v_1	PF00407.22	CL0209	Pathogenesis related protein Bet v 1 family	2	154	2.2E-17	1.1e-21
A0A7D9N3Y7	Bet_v_1	PF00407.22	CL0209	Pathogenesis related protein Bet v 1 family	2	154	4.5e-19	2.2e-23
A0A7D9N3Z9	Bet_v_1	PF00407.22	CL0209	Pathogenesis related protein Bet v 1 family			2.4e-18	1.2e-22
A0A7D9N402	Bet_v_1	PF00407.22	CL0209	Pathogenesis related protein Bet v 1 family	1	154	4.9e-22	2.5e-26
A0A7D9N413	Bet_v_1	PF00407.22	CL0209	Pathogenesis related protein Bet v 1 family	1	155	3.0e-19	1.6e-23

Table 7: HMMER result of 8 cashew protein allergens

Query	InterPro GO terms		
	Biological process	**Molecular function**	**Cellular component**
A0A1Z1G942	None	(GO:0048038) (GO:0051539) (GO:0008137) (GO:0051536)	None
A0A1Z1G958	(GO:0006412)	(GO:0003735)	(GO:0005840)
A0A7D9N3X7	(GO:0009738) (GO:0006952)	(GO:0038023) (GO:0010427) (GO:0004864)	None
A0A7D9N3Y2	(GO:0009738) (GO:0006952)	(GO:0038023) (GO:0010427) (GO:0004864)	None
A0A7D9N3Y7	(GO:0009738) (GO:0006952)	(GO:0038023) (GO:0010427) (GO:0004864)	None
A0A7D9N3Z9	(GO:0009738) (GO:0006952)	(GO:0038023) (GO:0010427) (GO:0004864)	None
A0A7D9N402	-	-	-

A0A7D9N413	(GO:0009738) (GO:0006952)	(GO:0038023) (GO:0010427) (GO:0004864)	None

Table 8: Gene ontology accession number of 8 cashew protein allergens from InterPro



Graph 1: Graph of 8 cashew protein allergens on the basis of gene ontology accession number on Web Gene Ontology Annotation Plot (WEGO) tool.

4.7 B-cell epitope prediction

To anticipate B cell epitopes of all 8 potential allergens, the Immune Epitope Database (IEDB) was used. The predicted epitopes and their locations are shown in table 9. The linear B-cell epitopes were predicted using the default parameters. Epitopes with lengths between 9 and 25 were chosen and used for further analysis.

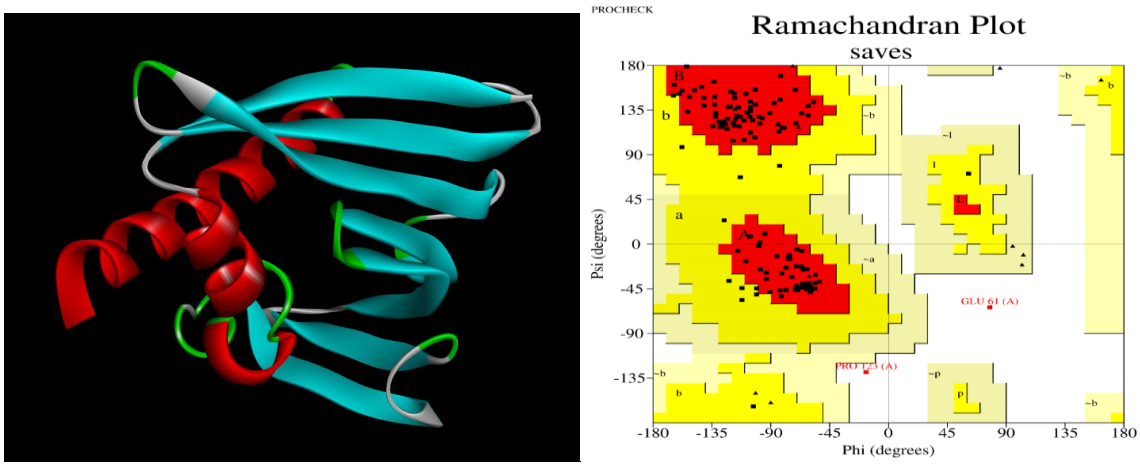
Query	Start	End	Length	Peptide
A0A1Z1G942	4	16	13	IEFSLRDRTTQNS
	117	128	12	TDSYSTVRGVDK
A0A1Z1G958	5	25	21	ISKRVLRRTGRIDSRKSARRI
	60	68	9	GFRGTRRGT
	119	135	17	VRDVTPMPHNGCRPPKK
A0A7D9N3X7	36	55	20	SQAIKSAELLQGDGGAGSIR
	71	79	9	DALDKETFV
	122	131	10	LPGFDVPGES
A0A7D9N3Y2	39	55	17	IKSAELLQGDGGAGSIR
	71	79	9	DALDKETFV
	122	131	10	LPGFDVPGES
A0A7D9N3Y7	41	53	13	AELISGDGGAGSI
	122	131	10	IVPGFEGAEN
A0A7D9N3Z9	41	55	15	SAELLQGDGGAGSIR
	71	79	9	DALDKETFV
	121	131	11	TLPGFDVPGES
A0A7D9N402	39	55	17	FKNIETIEGDGGPGTIK
	120	134	15	YYPKTGIELEEEKIK
A0A7D9N413	41	54	14	SAELISGDGGAGSI
	123	132	10	IVPGFEGAEN

Table 9: Total B-cell epitopes of screened 8 putative cashew allergens

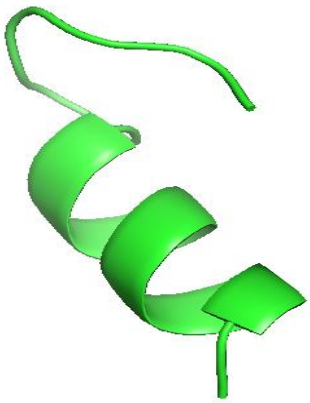
4.8 Structure Modelling

A Swiss model server was used to model all eight potential cashew allergens. The best model for the protein was selected for further analysis and Protein Data Bank (PDB) was referred to search the templates. Here, Figure 9 represent structure modelling data for cashew protein-Pathogenesis-related protein 10 **(A0A7D9N3Y2)**. Peptide structures were predicted by the Pep

Fold 3.5 server, which is shown in the figure 9.



A B



C

Figure 9: A) structure modelling data for cashew protein- Pathogenesis-related protein 10 **(A0A7D9N3Y2)** using Swiss-Model, **B)** Ramachandran plot of the cashew protein-Pathogenesis-related protein 10 **(A0A7D9N3Y2)** using Swiss-Model, and C) Peptide structures cashew protein- Pathogenesis-related protein 10 **(A0A7D9N3Y2),** peptide-IKSAELLQGDGGAGSIR predicted by the Pep Fold 3.5 server

4.9 Molecular Docking

A modelled structure was docked against an antibody structure obtained from the PDB. Docking was performed using the Antibody mode of the server Cluspro 2.0. Numerous docked models were produced by docking, but the model with the lowest binding energy has opted for additional analysis. Figure 10 shows the docked models for cashew protein- Pathogenesis-related protein 10 **(A0A7D9N3Y2)**, and he predicted linear conserved B-cell epitopes are

represented by rainbow-coloured spots in the figure. Similarly, table 10, displays the binding energy of the docked structure. While details about other cashew proteins are also depicted. Lowest binding energy of all the docking models was shown in table 11.

Cluster	Members	Representative	Weighted score
0	495	Center	-509.8
		Lowest energy	-852.7
1	184	Center	-534.1
		Lowest energy	-700.9
2	136	Center	-719.0
		Lowest energy	-719.0
3	128	Center	-511.3
		Lowest energy	-662.8

Table 10: Docking analysis of Pathogenesis-related protein 10 **(A0A7D9N3Y2)** using Cluspro 2.0 server.

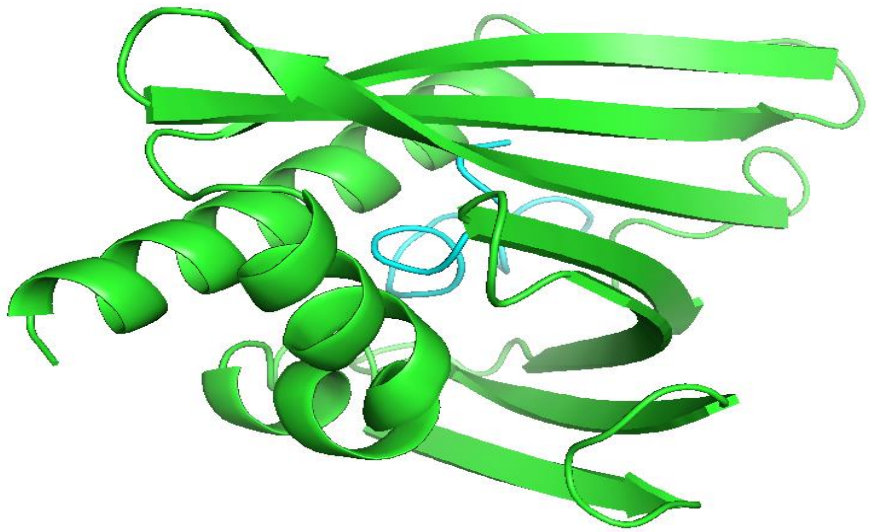


Figure 10 : Docked structure of Pathogenesis-related protein 10 **(A0A7D9N413)** using Cluspro 2.0 server

Number of Models	Lowest Binding Energy
A0A1Z1G942	**-509.1**
A0A1Z1G958	**-524.5**
A0A7D9N3X7	**-778.5**
A0A7D9N3Y2	**-852.7**
A0A7D9N3Y7	**-840.0**
A0A7D9N3Z9	**-772.2**
A0A7D9N402	**-877.2**
A0A7D9N413	**-764.2**

Table 11: Lowest binding energy of all the docking models

4.10 Expression analysis of identified allergens by using putative available transcriptomics data of *Anacardium occidentale.*

The gene sequencing expressing the allergen proteins was aligned against the transcriptomic data using BLAST, BLAST alignment resulted in the percent identity from 90-100% for all the query sequence implying that the allergen sequences aligned against the transcriptome have high gene expression (Table 12).

Allergen Protein sequences	Gene sequences
>tr|A0A1Z1G942|A0A1Z1G942_ANAOC NAD(P)H-quinone oxidoreductase subunit K, chloroplastic OS=Anacardium occidentale OX=171929 GN=ndhK PE=3 SV=1 MNSIEFSLRDRTTQNSVISTTSNDLSNWSRLSSLWPLLYGTSCCFIEFA SLIGSRFDFDRYGLVPRSSPRQADLILTAGTVTMKMAPSLVRLYEQM PEPKYVIAMGACTITGGMFSTDSYSTVRGVDKLIPVDVYLPGCPPKP EAVIDAITKLRKKISREIYEDRIRLQQENRSFPFTTNHKFRVVCSTPTG NYDQELLYQPPSTSEIPPETFFKYKSSVSSPEFVN	>NC_035235.1:c53591-52908 Anacardium occidentale chloroplast, complete genome ATGAATTCCATCGAGTTTTCCTTACGTGATCGAACAACCCAAAATTCAGTTATTTCAA CTACATCAAATGATCTTTCAAATTGGTCAAGACTCTCCAGTTTATGGCCCCTTCTTTA TGGTACGAGTTGTTGCTTCATTGAATTTGCTTCATTAATAGGCTCACGGTTCGACTTT GACCGTTATGGACTGGTACCAAGATCGAGTCCTAGACAGGCGGACCTAATTTTAACA GCTGGCACAGTAACAATGAAAATGGCTCCTTCTTTAGTGAGATTATATGAACAAATG CCTGAACCAAAATATGTTATTGCTATGGGAGCATGTACTATTACAGGGGGGATGTTC AGTACCGATTCTTATAGTACTGTTCGGGGAGTCGATAAGTTAATTCCGGTGGATGTTT ATTTGCCGGGATGTCCGCCTAAACCGGAGGCAGTTATAGATGCCATAACAAAACTTC GTAAGAAAATATCTCGAGAAATCTATGAAGATCGAATTCGATTGCAACAGGAGAATC GGTCGTTTCCTTTTACTACCAATCACAAGTTTCGTGTTGTATGCAGTACTCCTACTGG AAATTATGATCAAGAATTACTCTATCAACCACCATCTACGTCAGAAATCCCTCCGGA AACCTTTTTCAAATACAAAAGTTCAGTATCTTCCCCCGAATTCGTTAATTAG
>tr|A0A1Z1G958|A0A1Z1G958_ANAOC 30S ribosomal protein S11, chloroplastic OS=Anacardium occidentale OX=171929 GN=rps11 PE=3 SV=1 MAKSISKRVLRRTGRIDSRKSARRIPKGVIHVQASFNNTIVTITDVQG RVISWSSAGTCGFRGTRRGTPFAAQTAAGNAIRAVADQGMQRAEV MIKGPGLGRDAALRAIRRSGILLSFVRDVTPMPHNGCRPPKKRRV	>NC_035235.1:c82934-82518 Anacardium occidentale chloroplast, complete genome ATGGCAAAATCTATATCAAAAAGGGTTTTACGTAGGACTGGACGGATTGATTCACGT AAAAGTGCACGTCGAATACCAAAAGGCGTTATTCATGTTCAAGCAAGTTTCAACAAC ACCATTGTGACTATTACAGATGTACAGGGGCGGGTGATTTCTTGGTCCTCCGCTGGG ACTTGTGGATTCAGGGGTACAAGAAGAGGGACACCCTTTGCTGCTCAAACCGCAGCA GGAAATGCTATTCGAGCAGTAGCGGATCAAGGTATGCAACGAGCGGAGGTCATGAT AAAAGGTCCAGGTCTCGGAAGAGATGCAGCATTACGAGCTATTCGTAGAAGTGGTAT ACTTTTAAGTTTCGTACGGGATGTAACCCCTATGCCACATAATGGCTGCAGACCCCCT AAAAAAAGACGGGTGTAG

>tr|A0A7D9N3X7|A0A7D9N3X7_ANAOC Pathogenesis-related protein 10 OS=Anacardium occidentale OX=171929 PE=2 SV=1 MGFACGEFEIESVLPAAKMFQASVLDADQLFPKIMSQAIKSAELLQG DGGAGSIRKVKLV EGDSYMKHKVDALDKETFVYNYTIFEGDTLTDKFEKIVYETKWEST PAGGSIFKSSVKFY TLPGFDVPGESLLNKSKEKATAMVKAVEAYLQAN	>MN258367.1 Anacardium occidentale clone 18220-12 pathogenesis-related protein 10 mRNA, complete cds ATGGGATTTGCCTGCGGTGAATTTGAGATCGAGAGTGTCCTCCCAGCTGCCAAGATG TTCCAGGCCTCTGTCCTTGATGCTGATCAGCTCTTCCCCAAGATCATGTCTCAAGCTA TTAAGAGTGCCGAACTCCTTCAAGGTGATGGAGGGGCTGGAAGCATCAGGAAAGTT AAACTTGTTGAAGGCGACTCTTACATGAAGCACAAGGTTGATGCCCTGGACAAGGAA ACTTTCGTATACAACTACACCATCTTCGAAGGCGACACCTTGACTGACAAATTCGAG AAAATTGTTTACGAGACCAAGTGGGAGTCAACTCCTGCTGGAGGATCCATCTTCAAG TCCAGTGTCAAATTCTACACCTTACCCGGCTTCGACGTCCCCGGGGAGAGTTTACTTA ACAAGTCGAAGGAAAAGGCGACCGCCATGGTCAAGGCTGTTGAAGCATACCTCCAG GCCAATTAA
>tr|A0A7D9N3Y2|A0A7D9N3Y2_ANAOC Pathogenesis-related protein 10 OS=Anacardium occidentale OX=171929 PE=2 SV=1 MGFACGEFEIESVLPAAKMFQASVLDADQLFPKIMFQAIKSAELLQG DGGAGSIRKVKLVEGDSYMKHKVDALDKETFVYNYTIFEGDTLTDK FEKIVYETKWESTPAGGSIFKSSVKFYTLPGFDVPGESLLNKSKEKAT AMVKAVEAYLQAN	>MN258366.1 Anacardium occidentale clone 18220-11 pathogenesis-related protein 10 mRNA, complete cds ATGGGATTTGCCTGCGGTGAATTTGAGATCGAGAGTGTCCTCCCAGCTGCCAAGATG TTCCAGGCCTCTGTCCTTGATGCTGATCAGCTCTTCCCCAAGATCATGTTTCAAGCTA TTAAGAGTGCCGAACTCCTTCAAGGTGATGGAGGGGCTGGAAGCATCAGGAAAGTT AAACTTGTTGAAGGCGACTCTTACATGAAGCACAAGGTTGATGCCCTGGACAAGGAA ACTTTCGTATACAACTACACCATCTTCGAAGGCGACACCTTGACTGACAAATTCGAG AAAATTGTTTACGAGACCAAGTGGGAGTCAACTCCTGCTGGAGGATCCATCTTCAAG TCCAGTGTCAAATTCTACACCTTACCCGGCTTCGACGTCCCCGGGGAGAGTTTACTTA ACAAGTCGAAGGAAAAGGCGACCGCCATGGTCAAGGCTGTTGAAGCATACCTCCAG GCCAATTAA
>tr|A0A7D9N3Y7|A0A7D9N3Y7_ANAOC Pathogenesis-related protein 10	>MN258365.1 Anacardium occidentale clone 25514-15 pathogenesis-related protein 10 mRNA, partial cds

(Fragment) OS=Anacardium occidentale OX=171929 PE=2 SV=1 AVITDQLEVACTLPADKMFKGFVLDADDVFPKVMPQAIKSAELISG DGGAGSIRKVCVLEDDKLTYMKHKVDFLDRENLVFCYTIFEGDFLE SKFEKVVYETKWESGPDGGSIFKATAKFYIVPGFEGAENFITTEKEK AIGMIKAVEAHLKAN	GCAGTTATCACTGATCAACTGGAGGTGGCTTGTACTCTTCCTGCAGACAAGATGTTCA AGGGCTTCGTCCTTGACGCTGATGATGTCTTCCCCAAGGTTATGCCACAAGCTATTAA GAGTGCTGAACTCATCAGCGGCGATGGCGGAGCTGGAAGCATCAGGAAAGTTTGCG TTCTGGAAGACGACAAGCTCACTTACATGAAGCACAAGGTTGATTTTCTAGACAGAG AAAATTTGGTATTCTGCTACACCATCTTTGAAGGTGACTTTTTGGAAAGCAAGTTCGA GAAAGTTGTTTACGAGACCAAGTGGGAGTCAGGTCCTGATGGTGGATCCATCTTTAA GGCTACTGCCAAATTCTACATCGTACCTGGCTTTGAAGGCGCCGAGAATTTCATTACC ACCGAAAAAGAAAAGGCAATTGGCATGATCAAGGCTGTTGAAGCACACCTCAAGGC CAACTGA
>tr|A0A7D9N3Z9|A0A7D9N3Z9_ANAOC Pathogenesis-related protein 10 OS=Anacardium occidentale OX=171929 PE=2 SV=1 MGFACGEFEIESVLPAAKMFQASVLDADQLFPKIMSQAIKSAELLQG DGGAGSIRKVKLVEGDSYMKHKVDALDKETFVYNYTIFEGDTLTDK FEKIVYETKWESTPAGGSIFKSSVKFYTLPGFDVPGESLLNKSKEKVT AMVKAVEAYLQAN	>MN258368.1 Anacardium occidentale clone 18220-25 pathogenesis-related protein 10 mRNA, complete cds ATGGGATTTGCCTGCGGTGAATTTGAGATCGAGAGTGTCCTCCCAGCTGCCAAGATG TTCCAGGCCTCTGTCCTTGATGCTGATCAGCTCTTCCCCAAGATCATGTCTCAAGCTA TTAAGAGTGCCGAACTCCTTCAAGGTGATGGAGGGGCTGGAAGCATCAGGAAAGTT AAACTTGTTGAAGGCGACTCTTACATGAAGCACAAGGTTGATGCCCTGGACAAGGAA ACTTTCGTATACAACTACACCATCTTCGAAGGCGACACCTTGACTGACAAATTCGAG AAAATTGTTTACGAGACCAAGTGGGAGTCAACTCCTGCTGGAGGATCCATCTTCAAG TCCAGTGTCAAATTCTACACCTTACCCGGCTTCGACGTCCCCGGGGAGAGTTTACTTA ACAAGTCGAAGGAAAAGGTGACCGCCATGGTCAAGGCTGTTGAAGCATACCTCCAG GCCAATTAA
>tr|A0A7D9N402|A0A7D9N402_ANAOC Pathogenesis-related protein 10 OS=Anacardium occidentale OX=171929 PE=2 SV=1 MGVITFTEEFSSPVPARRLFKAFVLDFDNLLPKLMPQVFKNIETIEGD GGPGTIKKLNISEGGEVKYLKHRIDALDKEKLIYNYTIIEGDAMDKIE SVSYEIKYEVSPDGGCKGTTVNKYYPKTGIELEEEKIKEARAKAMGL	>MN258363.1 Anacardium occidentale clone 25355-15 pathogenesis-related protein 10 mRNA, complete cds ATGGGTGTCATCACTTTCACTGAAGAGTTTAGCAGCCCTGTCCCAGCTAGAAGATTGT TCAAAGCCTTCGTTCTTGATTTCGACAACCTCTTACCGAAGCTTATGCCTCAGGTTTTT AAGAACATCGAAACAATTGAAGGAGATGGAGGCCCTGGAACAATCAAGAAGTTGAA

YKVVEGYLLANPDAYA	CATCAGTGAAGGTGGGGAAGTTAAGTACTTGAAGCACAGGATTGATGCATTGGACA AAGAGAAGCTGATATACAATTATACAATAATTGAGGGCGATGCAATGGACAAGATT GAATCAGTTTCTTATGAGATTAAGTATGAGGTCTCCCCTGATGGAGGCTGCAAGGGT ACCACTGTTAACAAGTACTATCCGAAAACAGGCATCGAGCTTGAAGAAGAGAAAAT TAAGGAAGCCAGGGCAAAGGCCATGGGCCTCTATAAAGTTGTGGAAGGCTATCTCTT GGCAAATCCTGATGCCTATGCTTAA
>tr|A0A7D9N413|A0A7D9N413_ANAOC Pathogenesis-related protein 10 OS=Anacardium occidentale OX=171929 PE=2 SV=1 MAVITDQLEVACTLPADKMFKGFVLDADDVFPKVMPQAIKSAELIS GDGGAGSIRKVCVLEDDKLTYMKHKVDFLDRENLVFCYTIFEGDFL ESKFEKVVYETKWESGPDGGSIFKATAKFYIVPGFEGAENFITTEKEK AIGMIKAVEAHLKAN	>MN258364.1 Anacardium occidentale clone 25514-14 pathogenesis-related protein 10 mRNA, complete cds ATGGCAGTTATCACTGATCAACTGGAGGTGGCTTGTACTCTTCCTGCAGACAAGATG TTCAAGGGCTTCGTCCTTGACGCTGATGATGTCTTCCCCAAGGTTATGCCACAAGCTA TTAAGAGTGCTGAACTCATCAGCGGCGATGGCGGAGCTGGAAGCATCAGGAAAGTTT GCGTTCTGGAAGACGACAAGCTCACTTACATGAAGCACAAGGTTGATTTTCTAGACA GAGAAAATTTGGTATTCTGCTACACCATCTTTGAAGGTGACTTTTTGGAAAGCAAGTT CGAGAAAGTTGTTTACGAGACCAAGTGGGAGTCAGGTCCTGATGGTGGATCCATCTT TAAGGCTACTGCCAAATTCTACATCGTACCTGGCTTTGAAGGCGCCGAGAATTTCATT ACCACCGAAAAAGAAAAGGCAATTGGCATGATCAAGGCTGTTGAAGCACACCTCAA GGCCAACTGA

Table 12: Gene sequencing expressing the Allergen protein sequences

CHAPTER 5 DISCUSSION

The allergic reaction is mainly activated by food ingredients and affects the immune system of the individual, which vary depending on the individual's age and eating habit. A food allergy is an immune response to a food protein that would be innocuous. Mainly, food allergies are associated with the IgE and such food allergies typically affect atopic people who are genetically susceptible to allergies and who have already become sensitized to the allergen (Sicherer, 2000; Sicherer et al., 2006). Antigen-specific Immunoglobulin E (IgE), is mainly secreted after the first exposure to an offending food (also known as the sensitization phase), and it binds to the basophils' and mast cells' surfaces. When the antigen-specific IgE, cross-link on the surface of basophils and mast cells, which results in the release of numerous chemical mediators and as a result produces an allergic response (also known as the elicitation phase). These mediators then result in the clinical symptoms that are often linked to food allergies, such as itching, rhinitis, hives, or gastrointestinal symptoms, as well as anaphylaxis and bronchoconstriction on rare occasions. Tree nut allergy is reportedly seen in up to 0.05% to 4.9% of individuals worldwide, according to research from a systematic review. Among all tree nuts, cashews are the third most produced in the world. Cashew nuts rank third in worldwide production among all tree nuts. Cashew (*Anacardium occidentale*) belongs to the Anacardiaceae family and covers 9 species of the genus *Anacardium*. Even though cashews are one of the most popular and eatable tree nuts, there is very less information about their cross-reactive proteins. Thus, testing for allergenic proteins in cashews is necessary to learn more about the allergies to cashews, which are on the rise (da Costa et al., 2013). In this study, we aimed to identify potential cashew allergens using a genome-wide identification technique. The Codex Alimentarius Guidelines developed the method for determining whether a protein is allergenic in 2003 and 2009 (Commission, 2003). It recommended using informatics tools to compare the query protein sequences with known allergens as the first stage in the pre-identification process.

The identification of prospective candidates for further analysis is made easier with the use of in silico technologies, which can give a preliminary indication of the allergenic potential. Thus, the main aim of this study is to screen and identify the putative allergens in cashew nuts (*Anacardium occidentale).* Thus, based on sequence identity, the putative allergens are filtered out. Therefore, Sequence-based Characterization of predicted allergens of cashew nuts

(*Anacardium occidentale)* has been done. After that in silico validation of allergenicity among predicted allergens of *Anacardium occidentale* was evaluated. In order to determine a protein's ability to cause allergies based on its structural composition, ILSI (International Life Sciences Institute) and IFBC (International Food Biotechnology Council) first used a sequence identity technique (Vashisht et al., 2023). For the purpose of determining the allergenicity of a novel protein, allergen databases act as a repository for allergenic protein sequences. According to earlier research, sequence alignment is helpful in identifying any potential cross-reactivity between the query protein and listed allergens in the database.

Using the Basic Local Alignment Search Tool (BLAST) or FAST-All (FASTA) based algorithms, it is frequently possible to compare the sequences of cashew proteins with those of known allergens. Although it is based on a comparatively limited number of known allergens, the current criterion of 35% sequence similarity over 80 amino acids is a conservative approach.

Allergen databases have limitations, such as a there is a small number of allergen records, despite being that these tools are useful for screening potential cross-reactivity. Therefore, using a single database enhances the probability that some of the known allergens will be avoided. Due to these two databases (Allermatch and Allergen Online) were assessed for each protein to address this issue, and the final result was obtained using the consensus approach (Radauer et al., 2017). As the results from BLAST and FASTA (full and 80 amino acid formats), indicates and showed that different cashew proteins had a significant degree of sequence identity with known food allergens. A consensus method was used to narrow down the list of proteins. Only 11 out of 30 proteins (NAD(P)H-quinone oxidoreductase subunit K, Photosystem II reaction centre protein K, 30S ribosomal protein S11, ATP-dependent Clp protease, 30S ribosomal protein S7, and Pathogenesis-related protein 10) demonstrated consensus in these two allergen databases and plausible BLAST alignment results (Table 1), predicting potential cross-reactivity with potent food allergens. NAD(P)H-quinone oxidoreductase subunit K, Photosystem II reaction centre protein K, 30S ribosomal protein S11, ATP-dependent Clp protease, 30S ribosomal protein S7, and Pathogenesis-related protein 10- exhibited sequence identity with Pru p 1, ONCKE, IAA1_HORVU, Mal d 1, Pru p 1.0201, Pru av 1.0201, Pru ar 1, Pru av 1.0202, and Pru p 1.0301 respectively, which are all known to be allergens.

Further, three tools (AlgPred, AllerTOP, and Allergen FP) were used to demonstrate the allergenicity of these 11 cashew protein sequences based on their physicochemical properties. As the result compiled only eight cashew proteins appeared to be allergenic, using a consensus

approach. These sequences also displayed a high percentage of identity with Allergen Representative Peptides (ARPs), a collection of 24 amino acid long peptides that exhibit the highest global alignment with allergen proteins which depicted that the ARPs are present in the query sequences. therefore, these proteins are having a probability to be possible allergens. Additionally, the Cross-reactivity of eight cashew proteins with food allergens, notably *Prunus persica, Oncorhynchus keta, Prunus avium, Prunus armeniaca, and Malus domestica*, was discovered using a structure and sequence homology-based approaches. To assess the degree of sequence identity between the query sequence and a known allergen, sequence alignment using the MUSCLE tool was also performed. All 8 cashew proteins showed 30.4%-66.8% conservation in sequence alignment data using the MUSCLE tool, showing considerable sequence identity/homology at their amino acid level. All of these findings highlighted the possibility of cross-reactivity between cashew proteins and known food allergens and offered evidence that cashew proteins may be allergy-inducing. Along with the sequence-based analysis, structure-based analysis was also evaluated, with the help of PHD, PSIPRED, and I-TASSER tools for the prediction of secondary and three-dimensional structures respectively. Similar component values between cashew protein and food allergens were discovered, pointing to the structural similarity between the two and the potential allergic properties of cashew protein. Similarities between the two proteins (cashew and food allergy) at the residue level were further supported by similarities in amino acid makeup. When the two 3D structures of the query and reference proteins are superimposed, an RMSD value of <2 Å indicates a close fit, which may also be connected to their structural similarity and cross-reactivity. This demonstrated a close fit with the structure of a known allergen, which suggested similarity between the two structures, potential cross-reactivity, and allergenicity of the cashew proteins. Selected allergen sequences are classified into their families and gene ontologies using the databases: Pfam and Interpro respectively (Singh, A et al., 2020). As the structure of these allergen proteins had not been resolved previously and therefore is not even available in the protein data bank (PDB), thus Swiss-model server is used to choose the templates and also used to model the structure of the selected allergen proteins. Afterwards, the quality of the modelled structures of allergen proteins was evaluated by the Swiss-model score as well as with the help of the PROCHECK tool (Singh et al., 2022). Now, B- cell epitopes for the selected allergen proteins were predicted, and laterally their structures were modelled by the pepFold 3.5 servers (Mishra et al., 2016). After that Docking was performed using the Cluspro 2.0 server to generate the model (Bonvin et al., 2006). Numerous docked models were produced by docking, but the model with the lowest binding energy has opted for further analysis. Finally,

A0A7D9N3Y2 (Pathogenesis-related protein 10) was selected as the best-docked complex. However, in vitro, confirmation is still required. This necessitates conducting further trials to confirm the data. Therefore, additional research must be conducted to support the cashew nut's definite allergenicity and clinically evident cross-reactivity reactivity.

Expression analysis of the 11-allergen protein sequence gave a better insight into the expression pattern of these allergen sequences. Alignment of DNA, RNA or protein sequence can be due to the functional or structural similarity. Alignment of the query sequence against the transcriptomic data of *Anacardium occidentale* implies a high similarity between the two.

CHAPTER 6 CONCLUSION

Cashew nuts rank third in worldwide production among all tree nuts. Therefore, cashew is present all over the world, and they need to identify the allergens and it consumable for all. This study established identification and structural insights of allergens present in cashew nuts using computational tools. Cross-reactivity of eight cashew proteins with food allergens, notably *Prunus persica, Oncorhynchus keta, Prunus avium, Prunus armeniaca, and Malus domestica*, was discovered using a structure and sequence homology-based approaches. Further, to estimate the allergenicity of these cashew proteins, physicochemical properties-based assessments have been analysed. The degree of conservancy that cashew proteins have with food allergens was demonstrated using the secondary structure comparison, MUSCLE tool for sequence alignment, and three-dimensional structure. Classification of the allergen sequence along with their gene ontology was done which was further subjected to B- cell epitope prediction, structure modelling, and molecular docking. As a result, it is possible to assess the allergenicity of cashew-protein using an established methodology.

CHAPTER 7 REFERENCES

- Afzaal, M., Saeed, F., Niaz, B., Raza, M. A., Hussain, M., Aslam, M., ... & Ansari, M. J. (2023). Allergens in Cereal Grains. *Cereal Grains: Composition, Nutritional Attributes, and Potential Applications*.
- Alasalvar, C., & Shahidi, F. (Eds.). (2008). *Tree nuts: composition, phytochemicals, and health effects*. CRC press.
- Bahrami, A. A., Payandeh, Z., Khalili, S., Zakeri, A., & Bandehpour, M. (2019). Immunoinformatics: in silico approaches and computational design of a multi-epitope, immunogenic protein. *International reviews of immunology*, *38*(6), 307-322.
- Ben Rejeb, S., Abbott, M., Davies, D., Cléroux, C., & Delahaut, P. (2005). Multi-allergen screening immunoassay for the detection of protein markers of peanut and four tree nuts in chocolate. *Food additives and contaminants*, *22*(8), 709-715.
- Bignardi, C., Elviri, L., Penna, A., Careri, M., & Mangia, A. (2010). Particle-packed column versus silica-based monolithic column for liquid chromatography–electrospray-linear ion trap-tandem mass spectrometry multiallergen trace analysis in foods. *Journal of Chromatography A*, *1217*(48), 7579-7585.
- Blazowski, L., Majak, P., Kurzawa, R., Kuna, P., & Jerzynska, J. (2019). Food allergy endotype with high risk of severe anaphylaxis in children—Monosensitization to cashew 2S albumin Ana o 3. *Allergy*, *74*(10), 1945-1955.
- Bonvin, A. M. (2006). Flexible protein–protein docking. *Current opinion in structural biology*, *16*(2), 194-200.
- Borres, M. P., Sato, S., & Ebisawa, M. (2022). Recent advances in diagnosing and managing nut allergies with focus on hazelnuts, walnuts, and cashew nuts. *World Allergy Organization Journal*, *15*(4), 100641.
- Borres, M. P., Sato, S., & Ebisawa, M. (2022). Recent advances in diagnosing and managing nut allergies with focus on hazelnuts, walnuts, and cashew nuts. *World Allergy Organization Journal*, *15*(4), 100641.
- Borres, M. P., Sato, S., & Ebisawa, M. (2022). Recent advances in diagnosing and managing nut allergies with focus on hazelnuts, walnuts, and cashew nuts. *World Allergy Organization Journal*, *15*(4), 1004.

- Boyce, J. A., Assa'ad, A., Burks, A. W., Jones, S. M., Sampson, H. A., Wood, R. A., ... & Schwaninger, J. M. (2011). Guidelines for the diagnosis and management of food allergy in the United States: summary of the NIAID-sponsored expert panel report. *Journal of the American Academy of Dermatology*, *64*(1), 175-192.
- Breiteneder, H., & Radauer, C. (2004). Classification of plant food allergens. *Journal of Allergy and clinical immunology*, *113*(5), 821-830.
- Brusic, V., Petrovsky, N., Gendel, S. M., Millot, M., Gigonzac, O., & Stelman, S. J. (2003). Computational tools for the study of allergens. *Allergy*, *58*(11), 1083-1092.
- Burks, A. W., Tang, M., Sicherer, S., Muraro, A., Eigenmann, P. A., Ebisawa, M., ... & Sampson, H. A. (2012). ICON: food allergy. *Journal of Allergy and Clinical Immunology*, *129*(4), 906-920.
- Chen, H., & Wu, Y. (2022). Risk Assessment of Food Allergens. *China CDC Weekly*, *4*(34), 771-774.
- Crespo, J. F., James, J. M., Fernandez-Rodriguez, C., & Rodriguez, J. (2006). Food allergy: nuts and tree nuts. *British journal of nutrition*, *96*(S2), S95-S102.
- da Costa, J. S. B. (2013). *Development and application of molecular-based methods for the detection of tree nut allergens: the cases of almond, hazelnut and walnut* (Doctoral dissertation, Universidade do Porto (Portugal)).
- Datema, M. R., Van Ree, R., Asero, R., Barreales, L., Belohlavkova, S., de Blay, F., ... & Ballmer-Weber, B. (2018). Component-resolved diagnosis and beyond: multivariable regression models to predict severity of hazelnut allergy. *Allergy*, *73*(3), 549-559.
- Davis, L., Stonehouse, W., Loots, D. T., Mukuddem-Petersen, J., van der Westhuizen, F. H., Hanekom, S. M., & Jerling, J. C. (2007). The effects of high walnut and cashew nut diets on the antioxidant status of subjects with metabolic syndrome. *European journal of nutrition*, *46*, 155-164.
- Davoren, M., & Peake, J. (2005). Cashew nut allergy is associated with a high risk of anaphylaxis. *Archives of disease in childhood*, *90*(10), 1084-1085.
- Davoren, M., & Peake, J. (2005). Cashew nut allergy is associated with a high risk of anaphylaxis. *Archives of disease in childhood*, *90*(10), 1084-1085.
- De Leon, M. P., Glaspole, I. N., Drew, A. C., Rolland, J. M., O'Hehir, R. E., & Suphioglu, C. (2003). Immunological analysis of allergenic cross-reactivity between peanut and tree nuts. *Clinical & Experimental Allergy*, *33*(9), 1273-1280.

- Desta, I. T., Porter, K. A., Xia, B., Kozakov, D., & Vajda, S. (2020). Performance and its limits in rigid body protein-protein docking. *Structure*, *28*(9), 1071-1081.
- Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v. 2—a server for in silico prediction of allergens. *Journal of molecular modeling*, *20*, 1-6.
- Dimitrov, I., Naneva, L., Doytchinova, I., & Bangov, I. (2014). AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics*, *30*(6), 846-851.
- Elghoudi, A., & Narchi, H. (2022). Food allergy in children—the current status and the way forward. *World Journal of Clinical Pediatrics*, *11*(3), 253.
- Estruch, R., Ros, E., Salas-Salvadó, J., Covas, M. I., Corella, D., Arós, F., ... & Martínez-González, M. A. (2013). Primary prevention of cardiovascular disease with a Mediterranean diet. *New England Journal of Medicine*, *368*(14), 1279-1290.
- FAOSTAT, Food and Agriculture Organization of the United Nations, Statistics Division, Economic and Social Development Department, Rome, Italy. http://faostat3.fao.org/home/E. Accessed 12 Jan 2016.
- Fiers, M. W., Kleter, G. A., Nijland, H., Peijnenburg, A. A., Nap, J. P., & Van Ham, R. C. (2004). Allermatch™, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC bioinformatics*, *5*, 1-6.
- Fleischer, D. M., Conover-Walker, M. K., Matsui, E. C., & Wood, R. A. (2005). The natural history of tree nut allergy. *Journal of Allergy and Clinical Immunology*, *116*(5), 1087-1093.
- Gargano, D., Appanna, R., Santonicola, A., De Bartolomeis, F., Stellato, C., Cianferoni, A., ... & Iovino, P. (2021). Food allergy and intolerance: A narrative review on nutritional concerns. *Nutrients*, *13*(5), 1638.
- Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., van Ree, R., Vieths, S., & Taylor, S. L. (2016). AllergenOnline: a peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Molecular nutrition & food research*, *60*(5), 1183-1198.
- Grabenhenrich, L. B., Dölle, S., Moneret-Vautrin, A., Köhli, A., Lange, L., Spindler, T., ... & Worm, M. (2016). Anaphylaxis in children and adolescents: the European Anaphylaxis Registry. *Journal of allergy and clinical immunology*, *137*(4), 1128-1137.
- Gu, S., Chen, N., Zhou, Y., Zhao, C., Zhan, L., Qu, L., ... & Ding, Y. (2018). A rapid solid-phase extraction combined with liquid chromatography-tandem mass

spectrometry for simultaneous screening of multiple allergens in chocolates. *Food Control*, *84*, 89-96.

- Hasegawa, M., Inomata, N., Yamazaki, H., Morita, A., Kirino, M., & Ikezawa, Z. (2009). Clinical features of four cases with cashew nut allergy and cross-reactivity between cashew nut and pistachio. *Allergology International*, *58*(2), 209-215.
- Hefle, S. L., Nordlee, J. A., & Taylor, S. L. (1996). Allergenic foods. *Critical Reviews in Food Science & Nutrition*, *36*(S1), 69-89.
- Höfer, V., Martini, M., Dölle-Bierke, S., Worm, M., & Bilò, M. B. (2023). Health-related quality of life in food and venom induced anaphylaxis and role of influencing factors. *Clinical & Experimental Allergy*, *53*(3), 295-306.
- Johnson, J., Malinovschi, A., Alving, K., Lidholm, J., Borres, M. P., & Nordvall, L. (2014). Ten-year review reveals changing trends and severity of allergic reactions to nuts and other foods. *Acta paediatrica*, *103*(8), 862-867.
- Karnaneedi, S., Huerlimann, R., Johnston, E. B., Nugraha, R., Ruethers, T., Taki, A. C., ... & Lopata, A. L. (2020). Novel allergen discovery through comprehensive de novo transcriptomic analyses of five shrimp species. *International Journal of Molecular Sciences*, *22*(1), 32.
- Korte, R., Lepski, S., & Brockmeyer, J. (2016). Comprehensive peptide marker identification for the detection of multiple nut allergens using a non-targeted LC–HRMS multi-method. *Analytical and bioanalytical chemistry*, *408*, 3059-3069.
- Korte, R., Oberleitner, D., & Brockmeyer, J. (2019). Determination of food allergens by LC-MS: Impacts of sample preparation, food matrix, and thermal processing on peptide detectability and quantification. *Journal of proteomics*, *196*, 131-140.
- Lee, C. J., & McGill, S. K. (2023). Food allergies and alpha-gal syndrome for the gastroenterologist. *Current Gastroenterology Reports*, *25*(2), 21-30.
- Lee, L. A., & Burks, A. W. (2006). Food allergies: prevalence, molecular characterization, and treatment/prevention strategies. *Annu. Rev. Nutr.*, *26*, 539-565.
- López-Calleja, I. M., de la Cruz, S., González, I., García, T., & Martín, R. (2015). Development of real-time PCR assays to detect cashew (*Anacardium occidentale*) and macadamia (Macadamia intergrifolia) residues in market analysis of processed food products. *LWT-Food Science and Technology*, *62*(1), 233-241.
- López-Pedrouso, M., Lorenzo, J. M., Alché, J. D. D., Moreira, R., & Franco, D. (2023). Advanced Proteomic and Bioinformatic Tools for Predictive Analysis of Allergens in

Novel Foods. *Biology*, *12*(5), 714.

- Luparelli, A., Losito, I., De Angelis, E., Pilolli, R., Lambertini, F., & Monaci, L. (2022). Tree nuts and peanuts as a source of beneficial compounds and a threat for allergic consumers: Overview on methods for their detection in complex food products. *Foods*, *11*(5), 728.
- Mah, E., Schulz, J. A., Kaden, V. N., Lawless, A. L., Rotor, J., Mantilla, L. B., & Liska, D. J. (2017). Cashew consumption reduces total and LDL cholesterol: a randomized, crossover, controlled-feeding trial. *The American journal of clinical nutrition*, *105*(5), 1070-1078.
- Mah, E., Schulz, J. A., Kaden, V. N., Lawless, A. L., Rotor, J., Mantilla, L. B., & Liska, D. J. (2017). Cashew consumption reduces total and LDL cholesterol: a randomized, crossover, controlled-feeding trial. *The American journal of clinical nutrition*, *105*(5), 1070-1078.
- Mari, A., Scala, E., Palazzo, P., Ridolfi, S., Zennaro, D., & Carabella, G. (2006). Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. *Cellular immunology*, *244*(2), 97-100.
- Matricardi, P. M., Kleine-Tebbe, J., Hoffmann, H. J., Valenta, R., Hilger, C., Hofmaier, S., ... Ollert, M. (2016). EAACI Molecular Allergology User's Guide. Pediatric Allergy and Immunology, 27, 1–250.
- McWilliam, V., Koplin, J., Lodge, C., Tang, M., Dharmage, S., & Allen, K. (2015). The prevalence of tree nut allergy: a systematic review. *Current allergy and asthma reports*, *15*, 1-13.
- McWilliam, V., Koplin, J., Lodge, C., Tang, M., Dharmage, S., & Allen, K. (2015). The prevalence of tree nut allergy: a systematic review. *Current allergy and asthma reports*, *15*, 1-13.
- McWilliam, V., Koplin, J., Lodge, C., Tang, M., Dharmage, S., & Allen, K. (2015). The prevalence of tree nut allergy: a systematic review. *Current allergy and asthma reports*, *15*, 1-13.
- Medeiros-Linard, C. F. B., Andrade-da-Costa, B. L. D. S., Augusto, R. L., Sereniki, A., Trevisan, M. T. S., Perreira, R. D. C. R., ... & Lafayette, S. S. L. (2018). Anacardic acids from cashew nuts prevent behavioral changes and oxidative stress induced by rotenone in a rat model of Parkinson's disease. *Neurotoxicity Research*, *34*, 250-262.

- Mendes, C., Costa, J., Vicente, A. A., Oliveira, M. B. P., & Mafra, I. (2019). Cashew nut allergy: clinical relevance and allergen characterisation. *Clinical Reviews in Allergy & Immunology*, *57*, 1-22.

- Mendes, C., Costa, J., Vicente, A. A., Oliveira, M. B. P., & Mafra, I. (2019). Cashew nut allergy: clinical relevance and allergen characterisation. *Clinical Reviews in Allergy & Immunology*, *57*, 1-22.

- Messina, M., & Venter, C. (2020). Recent surveys on food allergy prevalence. *Nutrition Today*, *55*(1), 22-29.
- Metcalfe, D. D., Astwood, J. D., Townsend, R., Sampson, H. A., Taylor, S. L., & Fuchs, R. L. (1996). Assessment of the allergenic potential of foods derived from genetically engineered crop plants. *Critical Reviews in Food Science & Nutrition*, *36*(S1), 165-186.
- Mishra, A., Jain, A., & Arora, N. (2016). Mapping B-cell epitopes of major and minor peanut allergens and identifying residues contributing to IgE binding. *Journal of the Science of Food and Agriculture*, *96*(2), 539-547.
- Mohan, V., Gayathri, R., Jaacks, L. M., Lakshmipriya, N., Anjana, R. M., Spiegelman, D., ... & Willett, W. C. (2018). Cashew nut consumption increases HDL cholesterol and reduces systolic blood pressure in Asian Indians with type 2 diabetes: a 12-week randomized controlled trial. *The Journal of nutrition*, *148*(1), 63-69.
- Nagakura, K. I., Sato, S., Asaumi, T., Yanagida, N., & Ebisawa, M. (2020). Novel insights regarding anaphylaxis in children-With a focus on prevalence, diagnosis, and treatment. *Pediatric Allergy and Immunology*, *31*(8), 879-888.
- Nedyalkova, M., Vasighi, M., Azmoon, A., Naneva, L., & Simeonov, V. (2023). Sequence-Based Prediction of Plant Allergenic Proteins: Machine Learning Classification Approach. *ACS Omega*.
- New, L. S., Schreiber, A., Stahl-Zeng, J., & Liu, H. F. (2018). Simultaneous analysis of multiple allergens in food products by LC-MS/MS. *Journal of AOAC International*, *101*(1), 132-145.
- Nurmatov, U., Dhami, S., Arasi, S., Pajno, G. B., Fernandez-Rivas, M., Muraro, A., ... & Sheikh, A. (2017). Allergen immunotherapy for IgE-mediated food allergy: a systematic review and meta-analysis. *Allergy*, *72*(8), 1133-1147.
- Planque, M., Arnould, T., Renard, P., Delahaut, P., Dieu, M., & Gillard, N. (2017). Highlight on bottlenecks in food allergen analysis: Detection and quantification by mass spectrometry. *Journal of AOAC International*, *100*(4), 1126-1130.

- Pomés, A., Davies, J. M., Gadermaier, G., Hilger, C., Holzhauser, T., Lidholm, J., & Goodman, R. E. (2018). WHO/IUIS Allergen Nomenclature: Providing a common language. *Molecular immunology*, *100*, 3-13.
- Prado, M., Ortea, I., Vial, S., Rivas, J., Calo-Mata, P., & Barros-Velázquez, J. (2016). Advanced DNA-and protein-based methods for the detection and investigation of food allergens. *Critical reviews in food science and nutrition*, *56*(15), 2511-2542.
- Radauer, C. (2017). Navigating through the jungle of allergens: features and applications of allergen databases. *International archives of allergy and immunology*, *173*(1), 1-11.
- Radauer, C., Bublin, M., Wagner, S., Mari, A., & Breiteneder, H. (2008). Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *Journal of allergy and clinical immunology*, *121*(4), 847-852.
- Rico, R., Bulló, M., & Salas-Salvadó, J. (2016). Nutritional composition of raw fresh cashew (*Anacardium occidentale L.*) kernels from different origin. *Food science & nutrition*, *4*(2), 329-338.
- Robotham, J. M., Wang, F., Seamon, V., Teuber, S. S., Sathe, S. K., Sampson, H. A., ... & Roux, K. H. (2005). Ana o 3, an important cashew nut (*Anacardium occidentale L.*) allergen of the 2S albumin family. *Journal of allergy and clinical immunology*, *115*(6), 1284-1290.
- Ros, E. (2010). Health benefits of nut consumption. *Nutrients*, *2*(7), 652-682.
- Saha, S., & Raghava, G. P. S. (2006). AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic acids research*, *34*(suppl_2), W202-W209.
- Schein, C. H., Negi, S. S., & Braun, W. (2022). Still SDAPing Along: 20 Years of the Structural Database of Allergenic Proteins. *Frontiers in Allergy*, *3*, 863172.
- Sicherer, S. H., & Bock, S. A. (2006). An expanding evidence base provides food for thought to avoid indigestion in managing difficult dilemmas in food allergy. *Journal of allergy and clinical immunology*, *117*(6), 1419-1422.
- Sicherer, S. H., Furlong, T. J., Maes, H. H., Desnick, R. J., Sampson, H. A., & Gelb, B. D. (2000). Genetics of peanut allergy: a twin study. *Journal of Allergy and Clinical Immunology*, *106*(1), 53-56.
- Sicherer, S. H., Furlong, T. J., Muñoz-Furlong, A., Burks, A. W., & Sampson, H. A. (2001). A voluntary registry for peanut and tree nut allergy: characteristics of the first 5149 registrants. *Journal of Allergy and Clinical Immunology*, *108*(1), 128-132.

- Sicherer, S. H., Muñoz-Furlong, A., Godbold, J. H., & Sampson, H. A. (2010). US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. *Journal of Allergy and Clinical Immunology*, *125*(6), 1322-1326.
- Singh, A., & Upadhyay, A. K. (2022). Identification and annotation of peptide allergens in Prunus dulcis. *International Journal of Peptide Research and Therapeutics*, *28*(6), 1-9.
- Singh, A., Garg, S., & Upadhyay, A. K. (2023). Identification and analysis of allergens in edible mushroom (Agaricus bisporus). *Materials Today: Proceedings*.
- Tufail, T., Saeed, F., Ain, H. B. U., Niaz, B., Afzaal, M., Din, A., & Suleria, H. A. R. (2019). Cashew nut allergy; immune health challenge. *Trends in Food Science & Technology*, *86*, 209-216.
- Turnbull, J. L., Adams, H. N., & Gorard, D. A. (2015). The diagnosis and management of food allergy and food intolerances. *Alimentary pharmacology & therapeutics*, *41*(1), 3-25.
- Van Der Valk, J. P. M., J. Dubois, A. E., Gerth van Wijk, R., Wichers, H. J., & De Jong, N. W. (2014). Systematic review on cashew nut allergy. *Allergy*, *69*(6), 692-698.
- Vashisht, S., Singh, N., Sharma, A., Saini, N., Gaur, S. N., & Arora, N. (2023). In silico tools to assess the potential allergenicity of shiitake mushrooms (Lentinula edodes). *Journal of the Science of Food and Agriculture*, *103*(2), 877-890.
- Vashisht, S., Singh, N., Sharma, A., Saini, N., Gaur, S. N., & Arora, N. (2023). In silico tools to assess the potential allergenicity of shiitake mushrooms (Lentinula edodes). *Journal of the Science of Food and Agriculture*, *103*(2), 877-890.
- Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., ... & Peters, B. (2019). The immune epitope database (IEDB): 2018 update. *Nucleic acids research*, *47*(D1), D339-D343.
- Wang, C., Wang, Y., Liu, G., & Fu, L. (2020). Food allergomics based on high-throughput and bioinformatics technologies. *Food Research International*, *130*, 108942.
- Wang, F., Robotham, J. M., Teuber, S. S., Sathe, S. K., & Roux, K. H. (2003). Ana o 2, a major cashew (*Anacardium occidentale L.*) nut allergen of the legumin family. *International archives of allergy and immunology*, *132*(1), 27-39.

- Wang, F., Robotham, J. M., Teuber, S. S., Tawde, P., Sathe, S. K., & Roux, K. H. (2002). Ana o 1, a cashew (Anacardium occidental) allergen of the vicilin seed storage protein family. *Journal of Allergy and Clinical Immunology*, *110*(1), 160-166.
- Wang, L., Xiong, Q., Saelim, N., Wang, L., Nong, W., Wan, A. T. Y., ... & Tsui, S. K. W. (2023). Genome assembly and annotation of Periplaneta americana reveal a comprehensive cockroach allergen profile. *Allergy*, *78*(4), 1088-1103.
- Wei, Y., Sathe, S. K., Teuber, S. S., & Roux, K. H. (2002). A sensitive sandwich ELISA for the detection of trace amount of cashew nut in foods. *The Journal of Allergy and Clinical Immunology*, *1*(109), S303.
- Willison, L. N., Tawde, P., Robotham, J. M., Penney IV, R. M., Teuber, S. S., Sathe, S. K., & Roux, K. H. (2008). Pistachio vicilin, Pis v 3, is immunoglobulin E-reactive and cross-reacts with the homologous cashew allergen, Ana o 1. *Clinical & Experimental Allergy*, *38*(7), 1229-1238.
- Worm, M., Moneret-Vautrin, A., Scherer, K., Lang, R., Fernandez-Rivas, M., Cardona, V., ... & Grabenhenrich, L. B. (2014). First European data from the network of severe allergic reactions (NORA). *Allergy*, *69*(10), 1397-1404.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature methods*, *12*(1), 7-8.
- Ye , J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., ... & Shi, C. (2018). WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic acids research*, *46*(W1), W71-W75.
- Zhao, Y., Sun, X., Marquis, C. P., & Lee, N. A. (2019). Development of a sensitive sandwich ELISA specific to 2S albumin (Ana o 3) as a stable protein marker for cashew nut residue detection in pre-packaged food products. *Food Control*, *96*, 432-440.

PLAGIARISM REPORT

Document Information

Analyzed document	manuscript final thesis yadvi.docx (D172109773)
Submitted	2023-07-15 13:17:00
Submitted by	Atul Kumar Upadhyay
Submitter email	atul.upadhyay@thapar.edu
Similarity	2%
Analysis address	atul.upadhyay.thapar@analysis.urkund.com

Sources included in the report

SA	**2012-03-16_9650474.pdf** Document 2012-03-16_9650474.pdf (D5445286)	2
W	URL: http://www.allergen.org/ Fetched: 2023-07-15 13:17:00	3
SA	**ACS Omega.pdf** Document ACS Omega.pdf (D50056422)	1
SA	**JH_thesis akansha_plag draft.docx** Document JH_thesis akansha_plag draft.docx (D156863705)	1

Dr. Atul Kumar Upadhyay
Assistant Professor
Department of Biotechnology
Thapar Institute of Engineering and Technology
Patiala- 147004

Yadvi
602104019
M. Tech. Biotechnology
DBT, TIET
Patiala- 147004