Insights of amino acid usage patterns among allergenic and non-allergenic proteins

A Dissertation

Submitted to the Department of Biotechnology

Thapar Institute of Engineering and Technology

In partial fulfillment Of the Requirement for the Degree of

Master of Science in Biotechnology (July 2024)

Submitted by – Devvrat Prasad

M.Sc. Biotechnology (2022-24)

302201015

Supervisor – Dr. Atul Kumar Upadhyay & Dr. Vikas Handa

Assistant Professor

Department Of Biotechnology

Thapar Institute of Engineering and Technology

Thapar Institute of Engineering and Technology,

Patiala, Punjab

DECLARATION

I, **Devvrat Prasad**, certify that the thesis entitled "**Insights of amino acid usage patterns among allergenic and non-allergenic proteins**" is a record of original work I conducted under the supervision of **Dr. Atul Kumar Upadhyay**, Assistant Professor, and **Dr. Vikas Handa**, Assistant Professor, Department of Biotechnology, TIET. I further declare that this thesis has not previously formed the basis for awarding any degree, diploma, or other similar title to any candidate of this or any other university.

Devvrat Prasad
302201015

July 2024

CERTIFICATE

This is to certify that the dissertation report entitled "**Insights of amino acid usage patterns among allergenic and non-allergenic proteins**" submitted by **Devvrat Prasad** in partial fulfillment of the requirements for the award of the degree of **Master of Science in Biotechnology**, Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, is a record of his work carried out under my supervision and guidance. The report has not been submitted for any other degree or certificate award at this or any other university or institute.

Dr. Atul Kumar Upadhyay
Assistant professor
Department of Biotechnology
Thapar Institute of Engineering and Technology, Patiala

Dr. Vikas Handa
Assistant professor
Department of Biotechnology
Thapar Institute of Engineering and Technology, Patiala

ACKNOWLEDGMENTS

With profound gratitude, I especially acknowledge the exceptional guidance of Dr. Atul Kumar Upadhyay, **Assistant Professor**, Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, my esteemed project advisor. His unwavering encouragement and insightful supervision were a guiding light throughout this project. I am particularly grateful for the freedom he granted me to explore my ideas, deepening my passion for scientific research.

I also sincerely thank **Dr. Vikas Handa**, **Assistant Professor**, Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, my co-advisor. Dr. Handa's insightful supervision and constructive criticism were instrumental in ensuring the quality of my work. His dedication to my success is deeply appreciated.

My deepest appreciation extends to **Dr. M.S. Reddy**, **Head of the Department of Biotechnology**. His warm and approachable demeanor created a nurturing learning environment that significantly enriched my experience.

Sincere thanks to the entire department's non-teaching staff for their constant support. I am especially indebted to the PhD scholars, **Tanmayee Basu, Rashi Chugh, and Amardeep Singh**, whose expertise, willingness to share knowledge, and unwavering support were instrumental in overcoming various challenges. Their camaraderie was a constant source of encouragement.

Finally, to my dearest friends and family, my deepest gratitude. Your unwavering support, understanding, and belief in me were a source of immense strength throughout this project.

Thank you all from the bottom of my heart.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

RAAU	Relative amino acid usage
OFCs	Oral food challenges
FMT	Fecal microbiota transplantation
CRD	Chronic respiratory diseases
CoAn	Correspondence analyses

Abstract

Food allergies are a major public health problem that, at times, are seen to cause immunological cross-reactivity involving IgE and different families of allergens. This study investigates the amino acid usage patterns within four primary allergen families: Cupin, Bet v 1, Cap, and Tropomyosin. The aims of this work were to compare the structural and compositional features of allergenic and non-allergenic proteins based on several computational analysis, which included correspondence analysis using RAAU data, solubility analysis of the proteins, and well as Multiple Sequence Alignment.

Sequences of amino acids for allergens and non-allergens were obtained from public databases, and the sequences were compared. For the purpose of clustering the RAAU data, correspondence analysis (CoAn) was performed to find out the cluster pattern. When allergens were biplot against non-allergens within the allergen families from the CoAn biplots, it was evident that the two groups had significant qualitative variations in their amino acid signatures. Notably, Cupin allergens had differences in the distribution of the few amino acids viz., methionine, proline, and histidine, unlike the other groups, which were mostly characterized by the distribution of glutamine, glutamic acid, and tryptophan. Bet v 1 homolog allergens had a higher glycine content, affecting protein folding and possible allergenicity. However, tropomyosin allergens did not have a specific amino acid relationship that one could deduce, which needed more study.

The study also determined conserved motifs in each of the allergen families by multiple sequence alignments and comparisons that were at least 70% identical. It is assumed that these conserved motifs play a role in protein stability, allergenicity, and patterns of interactions with the immune system. This knowledge about conserved sequences can be valuable for the development of allergen-specific treatments and tests.

Solubility was determined by the principles of physicochemical properties using the SoluProt 1.0 tool. According to the assessment of the structural characteristics, the Cupin, CAP, and tropomyosin families of allergens displayed relatively low insolubility compared to non-allergenic proteins, while Bet v 1 allergens were relatively soluble. Similar differences may be postulated to affect the stability and interaction of these proteins with the immune system and thus their allergenic effects.

CHAPTER 1
INTRODUCTION

Food allergies are allergic reactions provoked by specific proteins contained in food sources referred to as allergens, which cause different reactions. These allergies severely affect global well-being and make the administrative aspects of medicine significantly more challenging (Sicherer & Sampson, 2014). Although several studies were conducted to find ways to address the issue, there is currently no known treatment for food allergies, so one is required to be very cautious about the foods one takes and always know how to handle a food allergy emergency, including the use of adrenaline.

The incidence of food allergies is rising globally, irrespective of the region developed or developing countries. A recently conducted study revealed that about forty-five percent of school-going children with food allergies ingest allergens by accident at least once a year, and most of these cases are relatively severe (mild to moderate) in nature (Ansotegui-Zubeldia & Fiocchi, 2023). The admission rates for food anaphylaxis range from 4 to 20 per 100,000 people, and the mortality rate has been reported to be much lower, approximately 0.03 per 100,000 (Ansotegui-Zubeldia & Fiocchi, 2023). In the general population, a food allergy is estimated to be present in every thousand, three hundred ninety-nine persons, and more specifically, for asthmatics, every three million, three hundred ninety-nine thousand nine hundred persons.

The management of food allergies must consider people's demography, level of awareness, ability to afford foods, and general food choices. The initial introduction of allergenic foods in weaning, such as peanuts, has shown some efficacy in avoiding food allergies (Peters et al., 2021). The mode of management is early diagnosis, proper treatment, including food immunotherapy and biologics like Omalizumab, and improved awareness of the symptoms. Recent approaches have emphasized patient-centered approaches, acknowledging that patients' needs and characteristics are different.

Food allergies are of three main types: immediate, delayed, and non-immediate. Type I involves the production of immunoglobulin E (IgE) the moment an individual is exposed to an

allergen, which causes immediate hypersensitivity such as rashes, itching, anaphylaxis, etc. (Francis et al., 2020). The delayed types of allergies include those with food-protein-induced enterocolitis syndrome and eosinophilic esophagitis, a disease that affects the T-cells, resulting in emesis, diarrhea, and abdominal pain after some time (Cocker et al., 2020). These types of reactions are more latent, less described, and difficult to address.

This has proven to be a significant challenge in both diagnosing and managing food allergies effectively. Cross-reactivity. The IgE antibodies formed against certain allergenic proteins can bind to proteins in other unrelated foods. For instance, people who are allergic to birch pollen may be affected by apples and peaches because of cross-reactive Bet v 1 homolog allergens (Francis et al., 2020). Awareness of these patterns enables the clinician to give patients the correct dietary information and minimize the limitations placed on them, such as restrictive diets, thereby improving their quality of life.

The study aims to look into IgE-mediated food allergies by looking at four main groups of allergens: cupin, the CAP family (Cysteine-Rich Secretory Protein), Bet v 1 homolog, and tropomyosin. Comparing structural and functional properties allows for explaining the reason for immune responses to some protein families. For instance, cupins in soybeans, to an extent, have a barrel shape and hence have an allergic effect on the bodies of many people.
Understanding each family's specific allergens and their occurrence in foods is critical to arriving at preventative measures. For instance, a fact regarding the heat-stable tropomyosin allergens present in shrimp preparation will help shellfish-allergic individuals avoid such foods. It aids in the development of allergen databases and the right food labels.

Thus, further investigation of the stability and solubility of the specific allergen families is needed. These properties determine the allergenicity and possibility of reactions. For example, the CAP family allergens of birch pollen and some fruits are dissimilar in their thermal stability and digestive effects. This information can be used to modify food processing techniques to minimize allergenicity, thereby enhancing the security of individuals who are allergic to certain foods.

Another concept, Relative Amino Acid Usage (RAAU), concerns the occurrence of specific amino acids in allergenic proteins. Research indicates that allergenic proteins have a higher

percentage of hydrophobic amino acids, making them more stable to degradative enzymes and, hence, having a higher allergenicity. Knowing RAAU offers awareness of the allergenicity of protein groups (Costa et al., 2021).

Descriptive correspondence analysis is one of the statistical techniques applicable to discrete attributes, such as amino acids in allergens. Researchers found differences in amino acid distribution; thus, they categorize proteins into allergenic and non-allergenic. Certain protein compositions may favor the formation of a structure or a phase where the proteins may well integrate, or the sequence could be less susceptible to enzymatic degradation, thereby enhancing its allergenicity (Radauer et al., 2011).

Nowadays, food allergies are considered to be a significant problem for global public health and need proper intervention. The allergens are not homogenous, and they share some characteristics with other substances, provoking an allergic reaction. Cross-reactivity creates difficulties in the diagnosis and treatment process. The purpose of this thesis is to investigate the allergenicity of cupin, CAP family, Bet v 1 homolog, and tropomyosin to improve knowledge of allergenicity. Greater understanding can result in improved diagnostic procedures, treatments, and prevention measures, as well as a higher quality of life for the food allergic patient. Understanding the families of allergenic proteins reduces the impact of allergies and presents a better life for patients.

CHAPTER 2
REVIEW OF LITERATURE

Food Allergy

Global acceptance has established food allergies as a major public health issue where the immune system reacts to food proteins. It manifests different clinical features with skin, respiratory, gastrointestinal, cardio, and neurological symptoms and signs (Anvari et al., 2019). This systematic review provides comprehensive data integration from various studies' examinations of food allergies, including etiology, diagnostic criteria, and treatment options.

Epidemiology and Prevalence of Food Allergies:

According to Kobelkova et al. (2021), food allergies are present worldwide, with 8% of children, 5% of adults, and even higher percentage rates in developed countries. Researchers also observed an increasing prevalence of food allergies, mainly in children in the developed world. This has been attributed to changes in diet and increased environmental factors, but it could also be blamed on the hygiene hypothesis, which states that reduced exposure to microbes in the first years of an individual's life predisposes such an individual to allergic diseases (Kobelkova et al., 2021).

Summarily, food allergies have also escalated in developing countries, although at a slower rate than those in developed countries. This trend depicts how individuals and organizations under globalization and Westernization regimes are surviving a recipe for food allergies in terms of diet and environment (Loh & Tang, 2018). Thus, according to the epidemiological information, there is a need to promote the establishment of international perspectives for the prevention and management of food allergies.

An observation made in Poland in the assessment of children's sensitization index revealed high sensitization to the most common food allergens among children; the allergens include peanuts, hazelnuts, and apples (Izabela et al., 2024). Comparisons of overall specific immunoglobulin E (sIgE) to these allergens among the different age groups were also

made, and this hinges on the features of sensitization that change with age. The data allows us to understand and find significance in the possibility of disaggregating predisposition towards allergens by region, or vice versa, depending on the effectiveness of potential intervention measures.

Pathophysiology of Food Allergies

Non-IgE-mediated food allergies are mainly categorized and consequently have distinct mechanisms and ways of occurrence. The food allergies cause an IgE-dependent immune response, which leads to a state of T-helper 2. These effects cause the antibodies, namely IgE, to firmly attach themselves to Fcε receptors present on effector cells such as the mast cell and basophils, thus triggering the release of histamine and other mediators to bring on the rapid onset of the symptoms (Anvari et al., 2019). These are the commonly displayed symptoms: urticaria, angioedema, and anaphylaxis, as well as respiratory troubles soon after the insect stings.

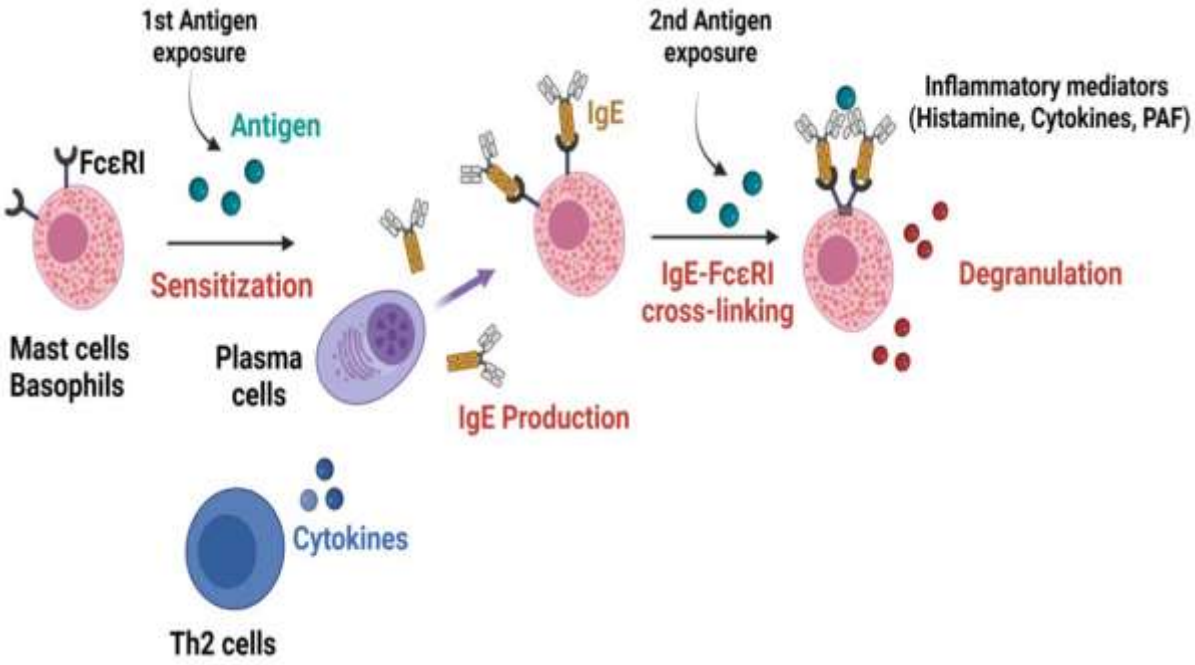


Figure 1- IgE-Induced Anaphylactic Response

However, it should be noted that the second type of food allergy, which is non-IgE, primarily affects the gastrointestinal tract. Secondary diseases like food protein-induced enterocolitis syndrome (FPIES) and allergic proctocolitis show symptoms later than other diseases. This is because the immune system is reacting, but the tissue in the GI tract that

is affected does not have pain receptors (Rojo-Gutiérrez et al., 2023). Some of the symptoms include vomiting and diarrhea, as well as the inability to gain weight normally or at all in babies. As a result, this study supports using Lactobacillus GG during exercise to alleviate symptoms by altering the growth of harmful bacteria, highlighting the significant role that microbiome modulation plays in managing food allergies that do not involve IgE antibodies.

Following these discoveries, which drew upon insights into the impact of the microbiota on immunity, numerous other valuable papers emerged. The researchers confirm the previously published findings, which show that patients with food allergies have different microorganisms colonizing their gastrointestinal tracts compared to those without the disorder. To treat this, microbiome-based therapies such as probiotics, prebiotics, and fecal microbiota transfer (FMT) have been put forward (Rojo-Gutiérrez et al., 2023) as possible choices.

Diagnostic Approaches

Diagnosing food allergies also encompasses the clinical history, specific IgE, and, at times, Oral food challenges (OFCs). In IgE-mediated allergies, separate tests are performed: skin-prick tests and serum-specific IgE tests, which are kinds of tests that are used to identify allergen-bound IgE (Tedner et al., 2022). The tests help in identifying the likely causative factors for allergies and the risk of allergic occurrences. That is, the severity of the reaction does not depend on the solidity of individual IgE or the maximum diameter of the skin test weal, while it is the onset of symptoms that may be related to such values.

When food allergies other than IgE are suspected, endoscopy and biopsy swabs have to be done to ascertain the level of inflammation and injury in the GIT. For instance, Eosinophilic oesophagitis (EoE), which is a condition related to food allergies, requires endoscopy to ascertain the presence of eosinophils in the oesophagus. A novel treatment that has also been employed for diagnostic purposes in the change of the intestinal microbiota in non-IgE-mediated food allergies includes fecal microbiota transplantation (FMT) (Rojo-Gutiérrez et al., 2023).

Diagnostic methods like Chronic respiratory diseases (CRD) have enhanced the opportunity for diagnosing food allergies by improving the technique used. Thus, CRD

helps pinpoint the allergenic proteins contained in foods, which in turn amplify the overall evaluation of a discrete subject's sensitization bracket. This aids in distinguishing between a true food allergy and the pollen-food syndrome, thereby guiding the appropriate management strategies (Kamath et al., 2023).

Management and Treatment Strategies

Food allergy management mostly focuses on the avoidance of certain foods that trigger allergy episodes and the correct management of any allergy episode. Epinephrine administered through intramuscular injection is considered the initial management of anaphylaxis, the most exemplary and fatal form of IgE-mediated food allergy (Anvari et al., 2019). Patients and their families need to be given information regarding the need to always be alert and the proper application of prescribed emergency drugs. Storable epinephrine auto-injectors should be easily accessible for any person at risk of anaphylaxis.

Some of the new guidelines even encourage early feeding of probable allergenic foods like peanuts to infants in a bid to progressively expose an infant to some of the foods that would have otherwise caused food allergies. Trials such as the LEAP (Learning Early About Peanut Allergy) have supported the notion that early introduction of peanut products to high-risk children, that is, children with severe eczema or egg allergies, lowers the probability of developing peanut allergies (Scarpone et al., 2023). Likewise, the EAT (Enquiring About Tolerance) study showed that the early introduction of multiple allergenic foods decreases the development of allergic diseases. However, this approach may also increase the withdrawal rates from the interventions due to the side effects experienced.

Allergen immunotherapy, or AIT, is another treatment for people with certain already well-identified food allergies. AIT is the process where gradually increasing concentrations of the allergen are introduced to the patient. Oral AIT (OIT), sublingual AIT (SLIT), and epicutaneous AIT (EPIT) also remain under research. The AIT is gradually being developed into the subcategories OIT, SLIT, and EPIT. These therapies seek to diminish the reactive capacity of the immune system about the allergen to lessen the seriousness of the allergic response (Anvari et al., 2019).

Newer Approaches and Directions

Current and future management of IgE-mediated food allergies, which include epicutaneous, sublingual, and oral immunotherapy, are regarded as clinical research. These approaches are meant to have the patient exposed to the allergen while steadily building up the dosage and reaction to it (Anvari et al., 2019). Furthermore, there are studies on IgG antibodies as a component of food allergies and natural tolerance. Lately, various IgG profiles have been described in persistent and transient FA, which may help understand mechanisms of spontaneous tolerance and design effective therapies.

There are monoclonal antibodies in development targeting particular components of the immune system with the optic to reduce allergic reactions, for instance, anti-IgE (omalizumab), anti-IL-4 receptor alpha (dupalimab), and anti-IL-5 (mepolizumab). Those biologic medicines are more precise in the treatment of food allergies and can be a replacement for immunotherapy procedures (Kamath et al., 2023).

Technologies like CRISPR-Cas9 seem to be the solution for food allergies soon. In order to achieve long-term effects that would eliminate the chance of an allergic reaction, researchers use genetic manipulation to remove or change genes that cause allergies. Though the area is relatively new for investigation, gene editing holds enormous promise for the treatment and control of food allergies and other allergic diseases (González-Daz, 2023).

Cross-reactivity and sensitivity patterns

The labeling of food allergens raises the problem of cross-reactivity, thereby complicating the diagnosis and treatment of those allergic to certain foods. In general, patients could have sensitization to many allergens belonging to the same food category, like tree nuts and peanuts, because the allergens have similar epitopes (Kamath et al., 2023). Given this, some cases may manifest clinical reactivity, while others may remain latent to symptoms. It may be concluded that the molecular explanation of cross-reactivity plays a critical role in both accurate diagnostics and suitable treatment plans.

A good example is the mono-sensitization to the peanut component Ara H6, while, though rare, its anaphylactic effects are severe. Identification of sensitization to the Ara h 6 peptide

is pertinent in patients with suspected peanut allergies where the presence of Ara h 3, Ara h 5, Ara h 7, and Ara h 8 is not present (van der Valk et al., 2016). Food allergens can additionally respond to other allergens, for example, airborne allergens comprising an equivalent protein molecule, for instance, pollen. This is called pollen-food syndrome or oral allergy syndrome, meaning that people with pollen allergies should not eat certain types of fruits, vegetables, or nuts.

The function of Maillard Reactions and Glycation End Products

The browning reaction that occurs during the heat treatment of foods leads to the generation of glycation end products that affect protein immunogenicity. These final products may increase the allergenic potential of food proteins, thereby deteriorating the state of patients with food allergies (Kobelkova et al., 2021). Other common effects of food sensitization include a decrease in endurance and performance in athletes, among other things. Further investigations are currently underway to either eliminate or lessen the effects of glycated end products on food allergies, as well as to evaluate which specific food preparation techniques can prevent the formation of these products.

There are probably many consequences of glycation end products and their involvement in different diseases, such as diabetes and cardiovascular diseases, whereas there is growing evidence for their connection with food allergies. Elucidation of the processes through which these products affect the allergenic potential might help in addressing the nature of food allergies and enhancing food security (Kobelkova et al., 2021).

Food allergy in special population

As such, there are specific characteristics of food allergy sensitization or prevalence among separate populations. For example, children of East Asian or African descent born in the Western world are more likely to be allergic to food compared to white children (Loh & Tang, 2018). This highlights the genome-environment interaction in food allergy development. It emerged that genetic makeup plays a role in people from such groups being more susceptible to food allergies through interaction with other factors like diet, lifestyle, and sensitization to food allergens.

Also, increased data highlights that not only the younger but also the elderly population suffer from food allergies, a factor that was not taken into account earlier. These are changes in environment and lifestyle, as well as the demographic factors of an aging population. The major offenders include peanuts, tree nuts, fish, and shellfish, for they are notorious for causing severe reactions among adults (González-Daz, 2023). Food allergies in the elderly are also likely to manifest in new and unusual symptoms and delayed reactions, which makes their diagnosis and treatment challenging at a certain age.

As a result, information on food allergies in pregnant women and their effects on the fetus and newborn is also important. Pregnant and breastfeeding mothers's diet and exposure to such allergens have a close connection with the development of food allergies in the progeny. However, the current literature remains unclear, necessitating further research to establish strict guidelines on food avoidance during pregnancy to prevent food allergies in this population (Loh & Tang, 2018).

Effect of Food Allergies on the Quality of Life

As already indicated, food allergies impact people's quality of life by altering dietary patterns, causing anxiety, and limiting those who have to adhere to certain diets. Concerning the reactions, one has to always be prepared for severe forms of intolerance, including anaphylactic reactions (Smolnikov et al., 2023). Patient education and their family enhancement form part of the management of the psychosocial aspect of food allergies. Special attention should be given to children who have food allergies because they encounter a lot of difficulties when they have to attend school and different parties. This results in social exclusion and can cause one to develop various forms of stress. Parents of the affected children also suffer stress and anxiety in relation to the future of their offspring in case they develop such allergies. Since the main problem in families with children with food allergies is a heavy emotional load, psychological support and counseling will be helpful here (Smolnikov et al., 2023). It is also critical for adults with food allergies to have organizations' accommodations and policies that allow them to work safely. The authorities should also be aware of food allergy risks and do everything to exclude the opportunity to meet an allergen accidentally, including the establishment of special allergen-free zones, clear marking of dishes, etc.

Therefore, it may be concluded that the rise in the incidences of food allergies across the globe highlights the importance of further investigations as well as the identification of efficient diagnostic and therapeutic tools. To improve knowledge about the treatments for food allergies, it is crucial to explore the mutual relationships among genetic, environmental, and immunological factors. Recommendations regarding the early introduction of allergenic foods, targeting individual patient profiles, and newly developed modalities open up rich possibilities for decreasing the impact of food allergies on individuals and increasing the quality of their lives.

Research Gap

Despite significant advancements in allergen research, there remain notable gaps in the understanding of specific gene families of allergens, particularly regarding their solubility and relative amino acid usage (RAAU). The literature on the solubility profiles of allergen gene families is sparse, with most studies focusing on broader allergenic properties rather than the detailed behavior of these proteins under various conditions.

This study focuses on four key allergen gene families: Cupin, Bet v 1, Cap family, and Tropomyosin. The Cupin family, known for its diverse group of major food allergens, was primarily studied for its allergenic properties, leaving solubility profiles under-explored. Bet v 1, a major birch pollen allergen also found in various fruits and vegetables, shows limited research on solubility and amino acid usage under different conditions. The Cap family, common in fruits like kiwi and papaya, is well-documented for enzymatic activities but lacks comprehensive data on solubility and RAAU. Tropomyosins recognized allergens in seafood and arthropods, are extensively studied for their allergenic roles but require detailed investigation into their solubility and amino acid usage.

This study aims to address these gaps by analyzing the solubility profiles and relative amino acid usage of these allergen families. Using advanced solubility prediction tools, RAAU analysis, and correspondence analysis, this research seeks to enhance understanding of how solubility and amino acid composition impact the allergenic properties and stability of these proteins, ultimately contributing to better allergen management strategies and informing food processing practices to mitigate allergenic risks.

Chapter 3
Material and Methods

1. Data Retrieval

Allergen protein sequences were retrieved directly from the UniProt Knowledgebase, a comprehensive public database curated by UniProt. UniProtKB offered immense data about the proteins, such as their sequences and functional descriptions. For more specific database searches, specific keywords related to the allergen families were used on the UniProtKB search page.

After retrieval, a methodical filtering process was performed carefully to achieve fine-quality sequence information with relevance to the research. In this case, entries were prioritized based on ‘expertly reviewed’ or ‘automatically annotated’ for data quality. Furthermore, the sequences started with food items commonly associated with allergies. Notably, only the protein entry that is considered an allergen in UniProtKB was included. Last, the obtained protein sequences were saved in FASTA format, which is a kind of textual representation format for protein data sequences.

Specifically, separate retrieval was done using nucleotide sequences that coded for the protein sequences. Nucleotide sequences for the allergen families were obtained from GenBank later on. The ID of the protein sequence from UniProtKB had been used as a query in BLAST or a similar homologous search from NCBI GenBank.

Nucleotide sequences were retrieved from GenBank (NCBI) for several allergen families. Specifically, 35 allergen and 18 non-allergen nucleotide sequences of the Cupin allergen family, 65 allergen and 23 non-allergen nucleotide sequences of the Bet v 1 family, 44 allergen and 22 non-allergen nucleotide sequences of the Cap family, and 30 allergen and 23 non-allergen nucleotide sequences of the Tropomyosin family had been obtained. These sequences comprised the final dataset.

Protein sequences were retrieved from UniProtKB. The final dataset for protein sequences included 48 allergen and 51 non-allergen sequences of the Cupin family, 82 allergen and 82 non-allergen sequences of the Bet v 1 family, 98 allergen and 74 non-allergen sequences

of the Cap family, and 35 allergen and 35 non-allergen sequences of the Tropomyosin family.

Specific websites and tools are employed for analyses such as RAAU analysis, correspondence analysis, and solubility analysis. These websites and tools, along with their descriptions and URLs, are listed in Table 1.

	NAME	DESCRIPTION	URL
1	UniProt-KB	UniProt-KB is the world's leading protein database, providing comprehensive and accurate information on their sequence, function, and annotation, with expertly reviewed (Swiss-Prot) and computationally predicted (TrEMBL) entries.	https://www.uniprot.org/uniprotkb/
2	PDB	The Protein Data Bank (PDB) is an online archive of 3D structures of proteins and nucleic acids, vital for understanding their function and designing new drugs.	https://www.rcsb.org/
3	NCBI	The National Center for Biotechnology Information (NCBI) acts as a digital library, curating a massive collection of genetic, protein, and medical research data freely accessible to scientists, students, and the public.	https://www.ncbi.nlm.nih.gov/
4	codonW	CodonW is a software tool used for analyzing codon usage patterns in DNA sequences, providing insights into evolutionary dynamics, gene expression preferences, and potential biases in protein translation.	https://codonw.sourceforge.net/
5	Genbank	GenBank is a comprehensive database containing annotated genetic sequences and associated metadata, freely accessible for scientific research and analysis.	https://www.ncbi.nlm.nih.gov/genbank/
6	Allergome	Allergome.org is an online database providing comprehensive information on allergens, including sequences,	https://www.allergome.org/

		structures, and clinical data, for research and clinical purposes.	
7	Soluprot 1.0	SoluProt is a web tool designed for predicting protein solubility, aiding in biotechnological and biochemical research applications.	https://loschmidt.chemi.muni.cz/soluprot/
8	Mega 11	MEGA 11 is a comprehensive software platform for molecular evolutionary analysis, phylogenetic tree construction, and the identification of conserved sites in genetic sequences, supporting advanced comparative genomics research.	**https://www.megasoftware.net/**

Table 1- Tools and websites used for Relative amino acid usage (RAAU) and Correspondence analysis.

2. Relative Amino Acid Usage or RAAU

To determine the relative abundance of amino acids in protein sequences, the Relative Amino Acid Usage (RAAU) was computed for each allergen family. This calculation shows how each amino acid is distributed among the protein's total amino acid content.

Therefore, there could be some related bias in amino acid composition within the allergen families that could be attributed to their allergenicity. The accession of sequences from GenBank (NCBI) was utilized in the sequence analysis of the sequences, which reflected the retrieved allergen protein sequences.

The queries concerning BLAST or similar homology applications in GenBank had been identified from UniProtK. In this step, although the strings themselves were no longer in nucleotide sequences in some cases, it must be noted that the information that resided in them became important for the subsequent process.

The following steps were performed on the retrieved nucleotide sequences:

Alignment: The obtained nucleotide sequences were analyzed using the right software program to establish the position or differences between the allergen gene families.

Amino Acid Composition Analysis:

i. These aligned nucleotide sequences were used in the prediction of the amino acids in the allergen proteins of each of the families. This involved identification of the amino acids that correspond with the required nucleotides for each of the sequences.

ii. RAAU Calculation: The quantity of each of the amino acids in each of the allergen families was calculated as a proportion of the total number of amino acids in the family and shown earlier with the help of the term RAAU.

iii. Furthermore, in this regard, the comparison of the quantitative values of the RAAU for each of the families was made with the help of the statistical test to identify significant differences in the patterns of the allergens and non-allergens, possibly related to the expression of allergenicity.

3. Correspondence Analysis (CoAn)

3.1 Data Preparation:

Selected allergen families were retrieved along with their protein sequences that were aligned in a common format.

Therefore, each protein sequence was translated into a binary matrix of an amino acid.

The binary presence-absence matrix created was then input into CodonW, a software that reads it in to begin the process of data analysis.

3.2 Performing Correspondence Analysis:

Sequence and amino acid usage were also normalized to 0–1 scale since the sequence length and the amino acid frequency may greatly differ.

The distance matrix for the rows (sequences) as well as the column (amino acids) was computed via CodonW using the chi-square distance matrix.

The Singular Value Decomposition (SVD) of the chi-square distance matrix was used to decompose the matrix into three components: known as the U Matrix, Σ (Sigma) Matrix, and V Matrix.

The diagonal elements of Σ are used to determine the proportion of variability, which represents the principal components of the amino acid usage differences in the sequences.

3.3 Biplot Visualization:

The principal components were used to perform a biplot, which is a two-dimensional plot that presents the relationship between sequences and amino acids.

The sequences and amino acids appeared as points in the biplot depending on their scores on the first two principal components.

3.4 Interpreting the biplot:

Clusters of Sequences: The sequences from different allergen families that positioned themselves close together on the biplot level would also be similar in terms of amino acid usage patterns.

Amino Acid Importance: Those amino acids, which map near specific families on the biplot, might play a crucial role in the action of this or those families.

Insights into Shared Features: As shown in Figure 3, the relative positions of these sequences and amino acids on the biplot to some extent revealed common or distinctive features of different allergen families regarding potential hidden features.

4. Solubility analysis

SoluProt 1.0 predicts protein solubility using the k-nearest neighbors algorithm, a machine-learning method. From this, the solubility of the proteins in gene allergen families is determined.

The analysis process involved several key steps:

Data Collection: Allergenic and nonallergenic protein sequences were retrieved from uniprotKB.

Data Preparation: To maintain the quality of the acquired dataset, further filtering was done to obtain sequences for redundancy and completeness.

Solubility Prediction: SoluProt 1.0 employed a machine learning algorithm that learned from large data sets of other soluble proteins. This algorithm used characteristics such as amino acid, sequence length, and physicochemical properties to determine solubility scores.

Compilation of Results: SoluProt 1 predicted the solubility scores obtained. Additionally, it compared the amino sequences gathered from allergenic proteins and non-allergenic proteins.

Thus, the use of a machine learning-based approach has helped in making precise predictions and significantly enhanced the knowledge of protein solubility and allergenicity.

5. Data Analysis

5.1 Result Analysis of RAAU and Correspondence Analysis

The Relative Amino Acid Usage (RAAU) data was obtained from codonW and subsequently used to plot a biplot graph. Statistically significant differences ($p < 0.05$) were revealed between the identified groups of sequences. These differences had been observed in the frequency with which certain amino acids were used within each family, both in allergen and non-allergen sequences.

5.2 Conserved Site at 70% Rate

Multiple sequence alignments using ClustalW were conducted. From this alignment, conserved motifs across all allergen families were identified at a 70% conservation rate.

5.3 Solubility Analysis of Allergens and Non-Allergens

The solubility of both allergen and non-allergen sequences was analyzed using SoluProt 1.0. Differences in solubility between these sequences were observed, providing insights into their structural and functional characteristics.

CHAPTER 4

RESULTS

1. Differential Amino Acid Usage Patterns among Allergen Gene Families:

This section focused on differences and similarities in the amino acid sequences of the cupin, CAP, Bet v 1 homolog, and tropomyosin allergens. The identified protein sequences had been subjected, for the purpose of the present research, to a specific parameter termed relative amino acid usage (RAAU) analysis, which was done using correspondence analysis (CoAn). Based on this study's results, its objective was to identify if there are developmental differences in the patterns of AA use within the sequences of each family and whether AA is an allergen or not.

1.1 Distinct Clustering Patterns Revealed by CoAn

It is important to mention that, by using the CoAn biplot shown in Figure 1, the sequences of allergens and non-allergens belonging to different gene families were clustered discretely. This meant that allergenic sequences had a different usage profile of amino acids compared to non-allergenic sequences. Further analysis of the biplot identified specific amino acid associations with each family. The further study of the biplot had allowed particular combinations of amino acids with each family to be known:

- Cupin allergens: opportunities for the addition of methionine, proline, and histidine were identified. This might be related to the seed storage and protein-stabilizing roles of such molecules within the cells.
- CAP allergens: A possible increase in the amino acid profile of the protein, especially of glutamine, glutamic acid, and tryptophan, had been noted. The effect of these amino acids on solubility and, thus, the possible interaction with immune cells might be different.
- Bet v 1 homolog allergens: It might also have been further enriched with glycine. Thus, this amino acid might change the processes of protein folding that could influence the development of allergenicity through increased resistance to proteolysis.

- Tropomyosin allergens: From the results analysis, tempting relations with the above-mentioned conditions of specific amino acids could not have been established. Further studies might be conducted in order to discover any other patterns that might be related to this family.

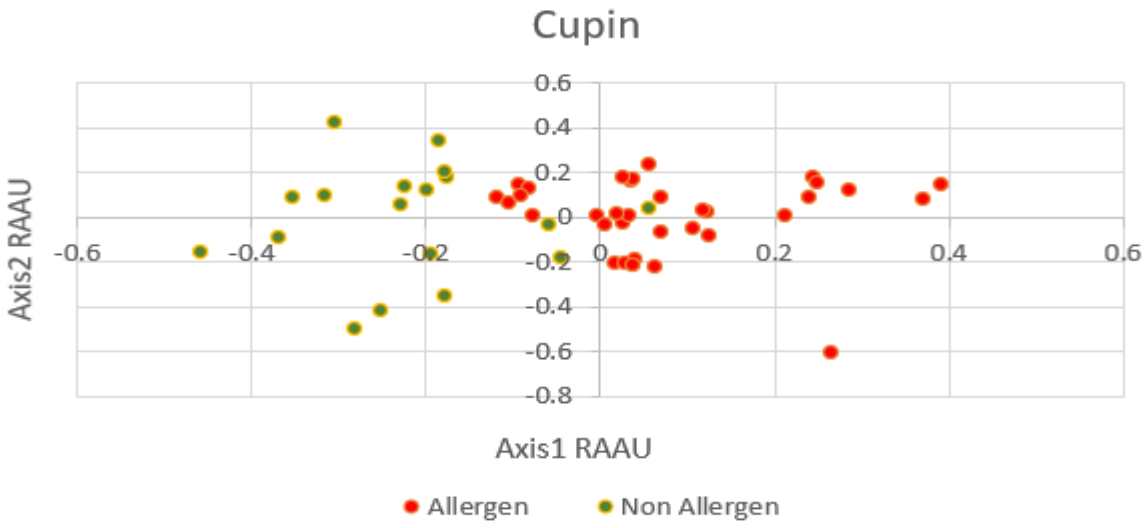


Figure 2- Correspondence analysis of the Cupin gene family plotted against Axis 1 and Axis 2 of RAAU data. The allergen genes are represented as red, whereas non-allergens are denoted as green.

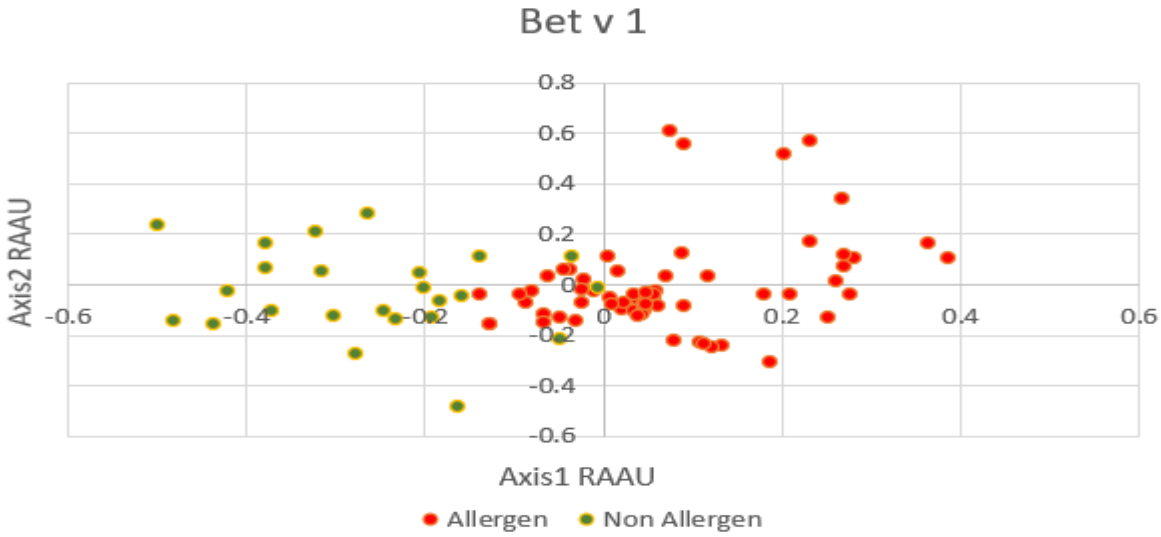


Figure 3- Correspondence analysis of the Bet v 1 gene family plotted against Axis 1 and Axis 2 of RAAU data. The allergen genes are represented as red, whereas non-allergens are denoted as green.

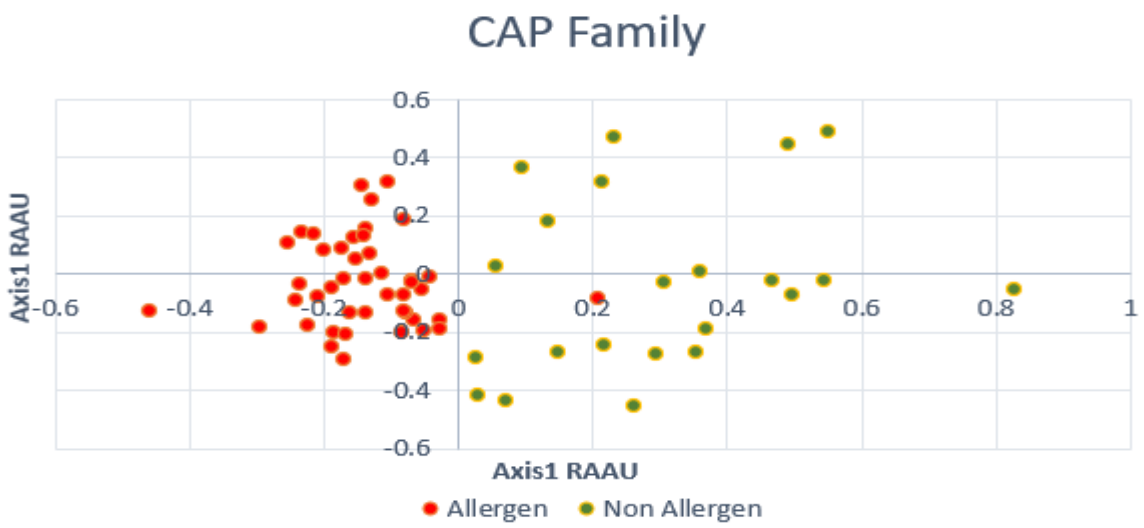


Figure 4- Correspondence analysis of the CAP Family gene family plotted against Axis 1 and Axis 2 of RAAU data. The allergen genes are represented as red, whereas non-allergens are denoted as green.

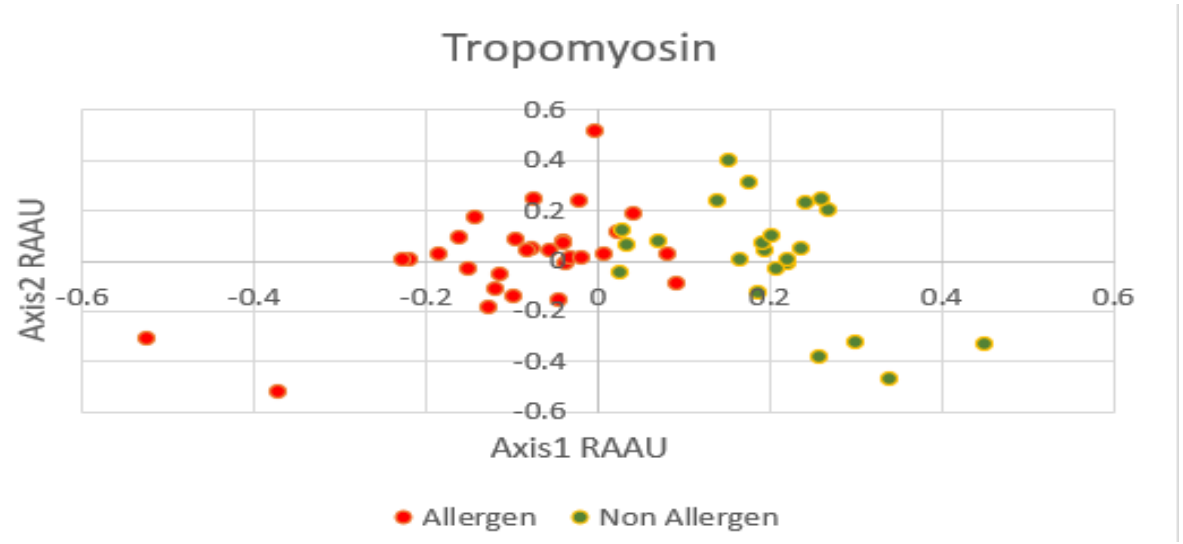


Figure 5- Correspondence analysis of the Tropomyosin gene family plotted against Axis 1 and Axis 2 of RAAU data. The allergen genes are represented as red, whereas non-allergens are denoted as green.

These findings showed that sequences within the respective gene families showed different patterns of amino acid usage, with allergens highlighted in red and non-allergens in green on the graphs. All the gene family graphs clearly showed distinct patterns of amino acid usage in their sequences.

1. 2 The variation of amino acid usage between allergen and non-allergen sequences

- As seen in Table 1, statistically significant differences between the identified groups of sequences ($p<0.05$) have been revealed in the frequency with which certain amino acids are used within each family, both in allergen and non-allergen sequences. Allergen sequences had exhibited a higher abundance of: Allergen sequences exhibited a higher abundance of:

- Methionine
- Proline
- Histidine
- Glutamine
- Glutamic acid
- Tryptophan
- Glycine is also found in Bet v 1 homologs.

- These findings provided the basis for the observations made from the CoAn biplot analysis and indicated that certain amino acids may have an association with allergenicity present in these families.

Amino acid (cupin)	RAAU allergen	RAAU non-allergen	Significance level (*P)*
Over-represented amino acids			
Glutamine	55	24.5	<0.05
Glutamic acid	64	39	<0.05
Serine	15.7	14.66	<0.05
Asparagine	37.5	28	<0.05
Lysine	24.5	15.5	<0.05
Arginine	13.375	12.6	<0.05
Isoleucine	24	12	<0.05
Under-represented amino acids			
Methionine	7	20	<0.05
proline	14	23.5	<0.05
Histidine	9	25	<0.05
Tryptophan	3	40	<0.05

Glycine	16.5	29	<0.05
Valine	16.33	23.75	<0.05
Threonine	8.25	13.33	<0.05
Alanine	10.5	25.66	<0.05
cysteine	0.75	15.5	<0.05
Phenylalanine	21	27	<0.05

Table 2- Over/under-represented amino acids based on normalized RAAU of the allergen and non-allergen cupin

Amino acid (CAP family)	RAAU allergen	RAAU non-allergen	Significance level (*P*)
Over-represented amino acids			
Methionine	16	22	<0.05
proline	11	5	<0.05
Histidine	16	11	<0.05
Glutamine	22	12	<0.05
Glutamic acid	24.5	11	<0.05
Tryptophan	17	10	<0.05
Glycine	17.5	3.75	<0.05
Valine	17.75	12	<0.05
Serine	10.5	9.16	<0.05
Threonine	15.25	8.25	<0.05
Alanine	22	4.25	<0.05
Asparagine	26.5	12	<0.05
Lysine	26.5	25.5	<0.05
Arginine	7.33	5	<0.05
cysteine	16	5.5	<0.05
Isoleucine	20.5	18.33	<0.05

Table 3- Over-represented amino acids based on normalized RAAU of the allergen and non-allergen CAP family

- Additional research had been necessary to describe the functional implications of these various amino acid utilization profiles and to explain how they could pave the way for immune reactions. But this initial analysis had given some idea about the relationship, if any, between amino acid composition and allergenicity within these allergen families.

2. Conserved Motifs Identified Within Allergen Gene Families

- The study identified conserved motifs in the allergenic sequences of the cupin, Bet v 1 homolog, CAP family, and tropomyosin cupin superfamily.
- Bioinformatics calculations were performed to construct multiple alignments of the allergen regions of the different families using MEGA 11 software.
- These alignments carried out by the ClustalW algorithm helped in pointing out areas where there were strong matches or even identities of the amino acid residues in the sequences.
- They were defined as subsequences identified in the alignments that had a minimum of 70% identity between the aligned sequences of that particular allergen family.

2.1 Sequence Alignments and Motif Identification

- The MEGA 11 source enabled the development of multiple sequence alignments for each allergen family; these alignments offered a view of the similarities and differences between the amino acid sequences.
- These alignments were necessary for selecting conserved motifs that could possibly be areas of functional or structural significance within the allergen proteins.
- Such a selection criterion of 70% identity level was set with the view of hitting considerable sequence homologies while at the same time minimizing the number of motifs for analysis.

conserved sequence site at 70% identity level			
CUPIN	CAP family	BET V 1	TROPOMYOSIN
GS			
EDVERLIKNQQQSYFANAQPQQQQQ	MIAFIVLPILAAVALQQSSG	FTYET	MD
REREGRHGRRG	VDFDSESPRKPEIQN	VIPPPR	KKKMQA
LRLLGFGINADE	DL	AFIL	DTL
QRNFLAGSEDNVIRQLDREVK	SL	NLI	QQNK
HY	RS	IAPQAV	EAN
SK	PT	KI	RAEK
EGNYEL	SNMLKME	EGSQFGSVTHKIDGI	DL
GE	YPEA	KISYETKLVASSD	QVQ
DIFVI	AN	STSNYHT	GEVAA
AGHPISIN	ERW	VEIKEEHV	QA

SSN	YR	AGKEKASHLFKLV	NRSLS
AQSSS	RVLG	LANPNEYC	MDA
SGPFNLRSNKPIYSNKF	IK		RFL
NFYEITP	CGENIYMSP		MV
TTSYILNPDDN	ADPPNA		
QNLRVVKLAIPINNPG	YKSY		
YDFYPSSTKD	AAAYCPSS		
	YSYFYV		

Table 4- Gene allergen families Conserved sequence site at 70% identity level

- Figures of CUPIN allergens conserved sequence site at 70% identity level.

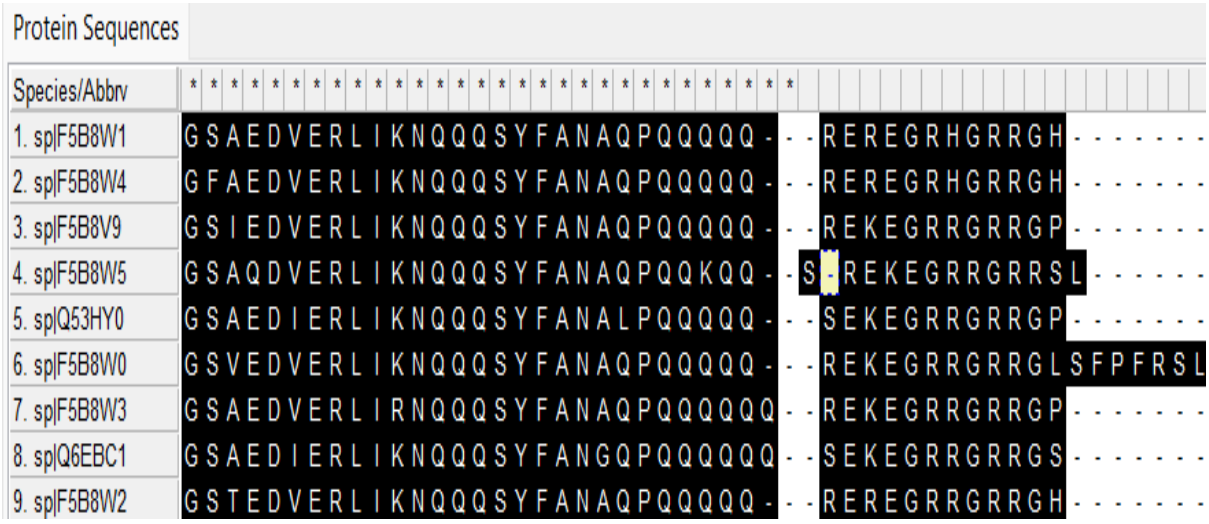


Figure 6 – CUPIN Conserved sequence site at 70% identity level (GS, EDVERLIKNQQQSYFANAQPQQQQQ, REREGRHGRRG)

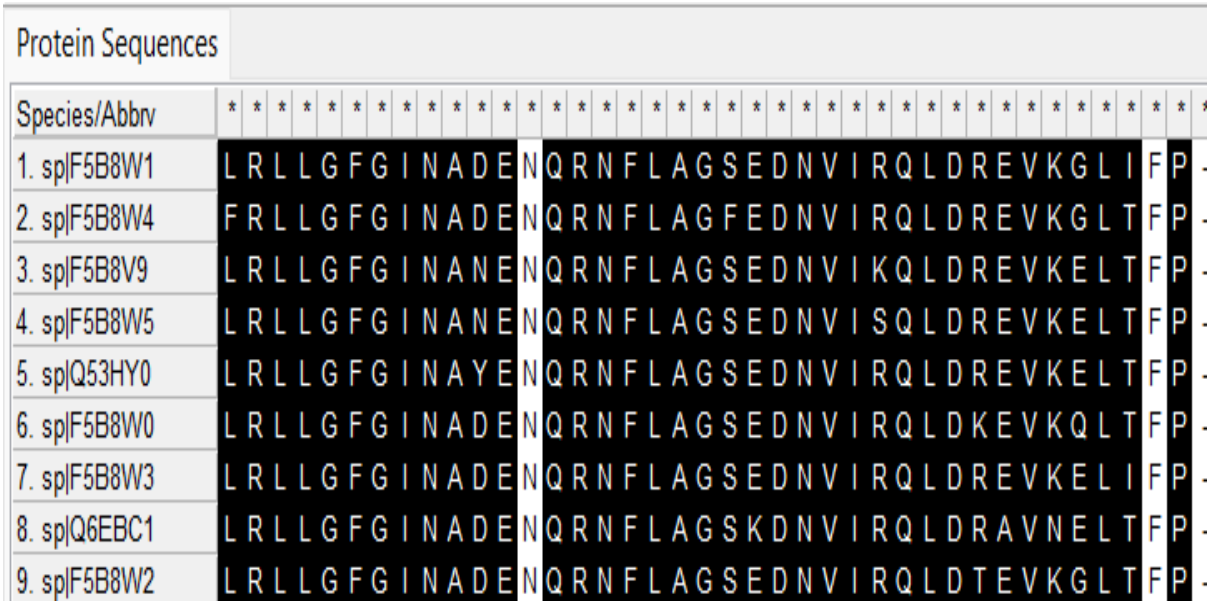


Figure 7 – CUPIN Conserved sequence site at 70% identity level (LRLLGFGINADE, QRNFLAGSEDNVIRQLDREVK)

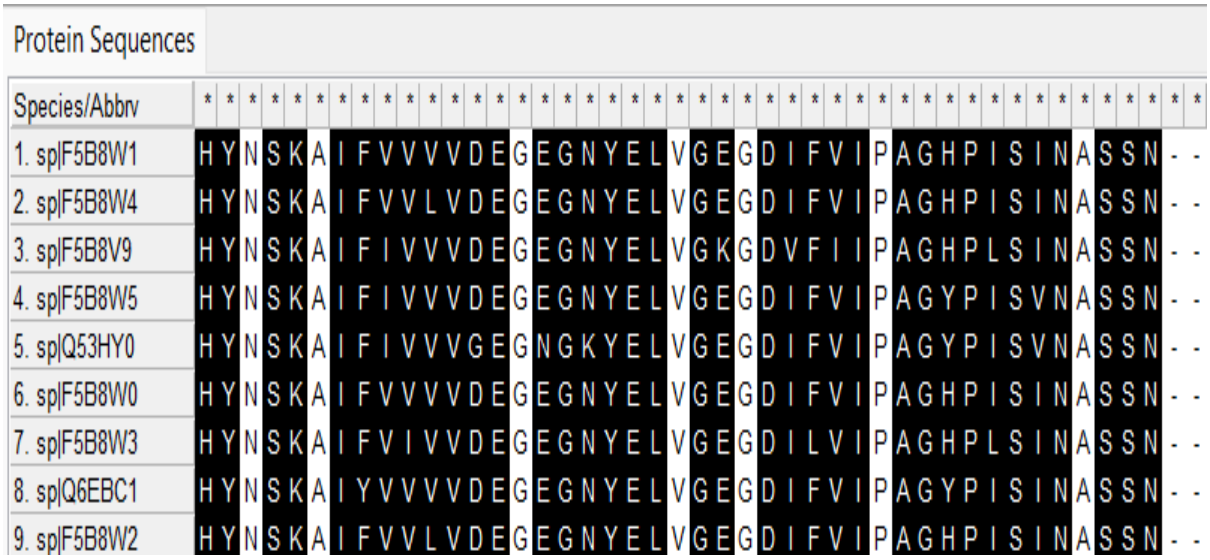


Figure 8 – CUPIN Conserved sequence site at 70% identity level (HY, SK, EGNYEL, GE, DIFVI, AGHPISIN, SSN)

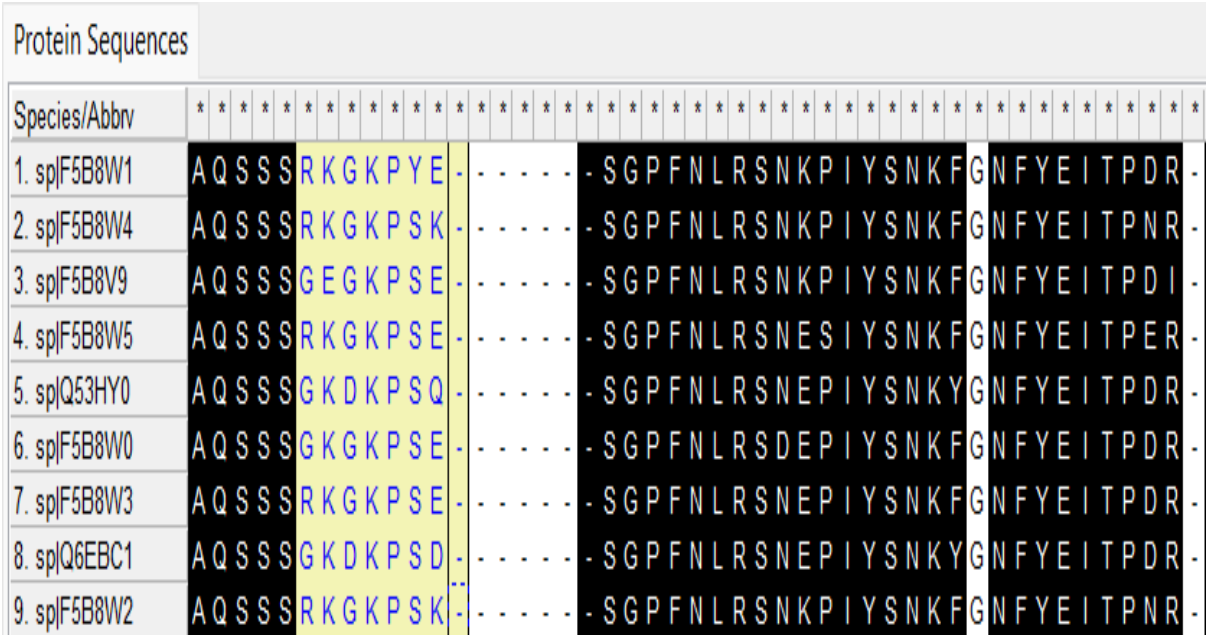


Figure 9 – CUPIN Conserved sequence site at 70% identity level (AQSSS, SGPFNLRSNKPIYSNKF, NFYEITP)

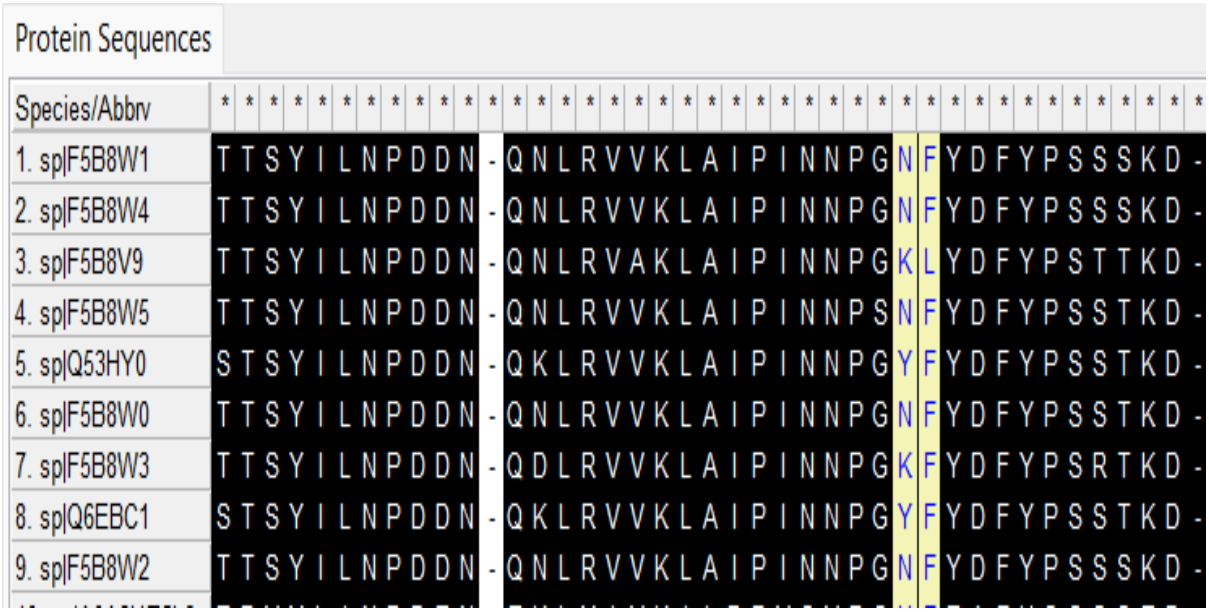


Figure 10 – CUPIN Conserved sequence site at 70% identity level (TTSYILNPDDN, QNLRVVKLAIPINNPG, YDFYPSSTKD)

- Figures of CAP Family allergens conserved sequence site at 70% identity level.

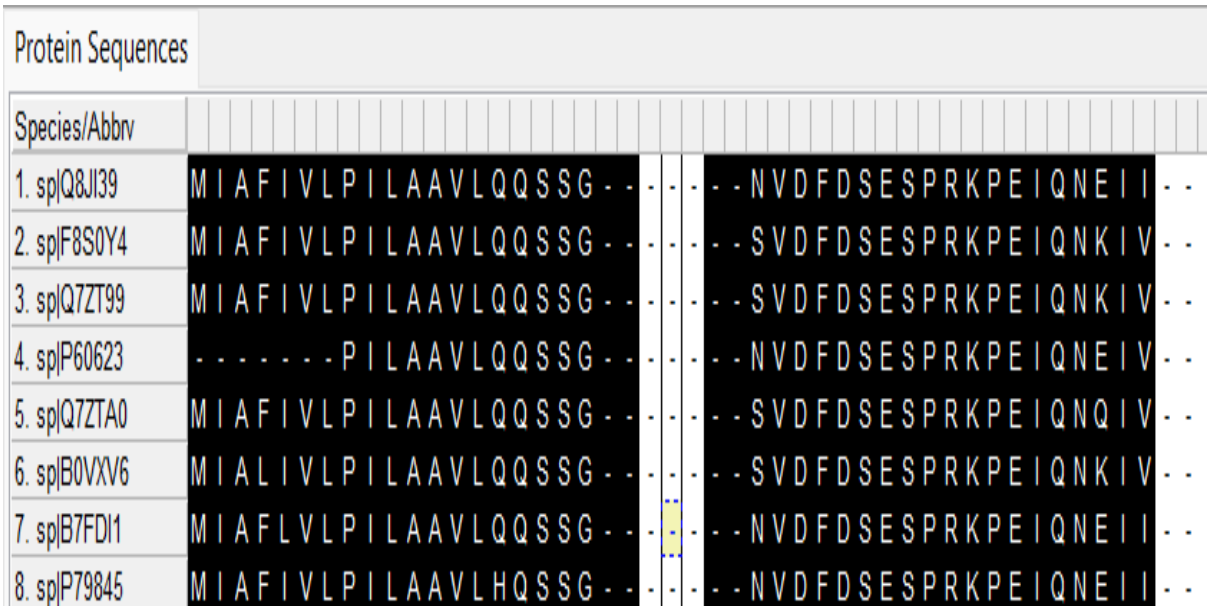


Figure 11 – CAP Family Conserved sequence site at 70% identity level (MIAFIVLPILAAVALQQSSG, VDFDSESPRKPEIQN)

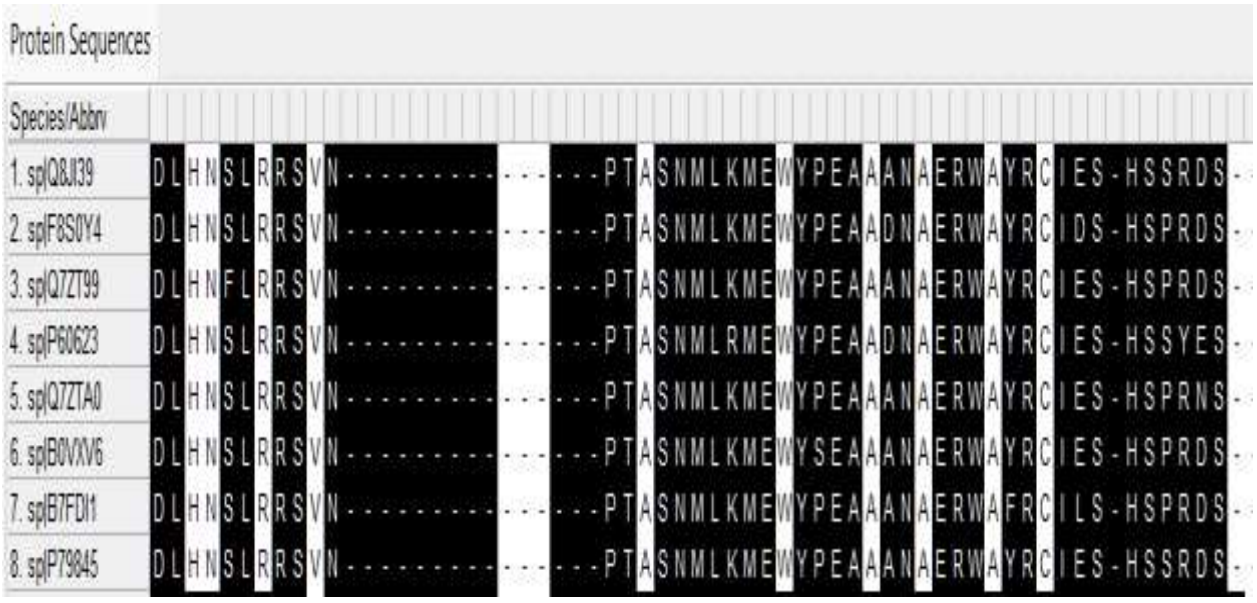


Figure 12 – CAP Family Conserved sequence site at 70% identity level (DL, SL, RS, PT, SNMLKME, YPEA, AN, ERW, YR)

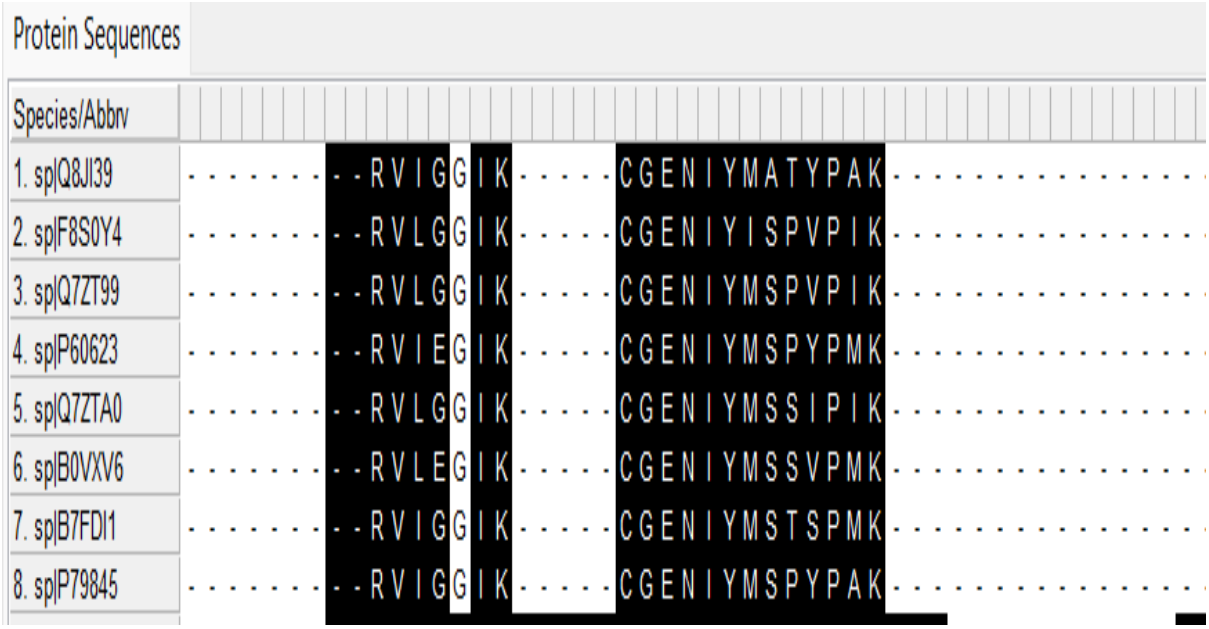


Figure 13 – CAP Family Conserved sequence site at 70% identity level (RVLG, IK, CGENIYMSP)

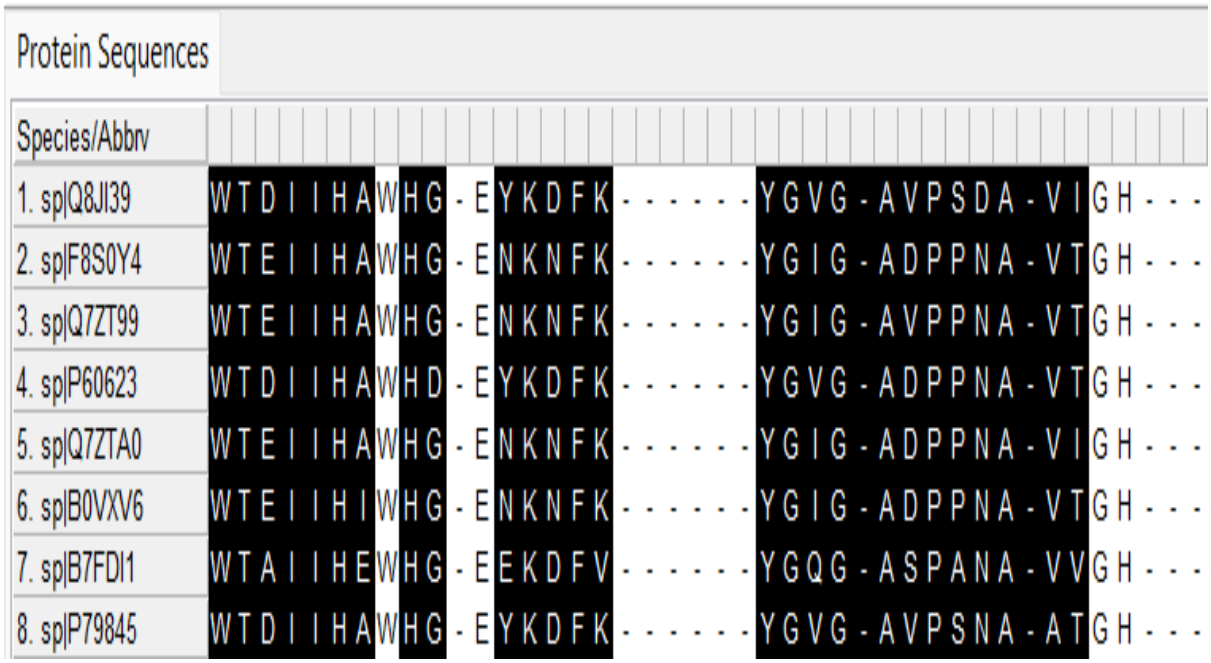


Figure 14 – CAP Family Conserved sequence site at 70% identity level (ADPPNA)

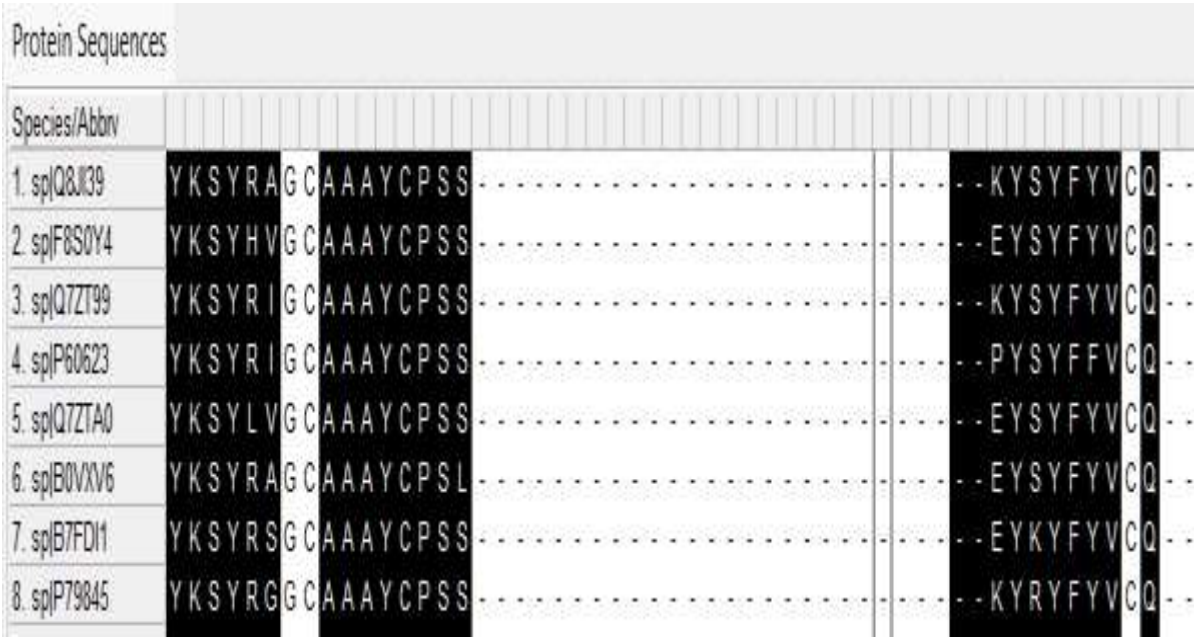


Figure 15 – CAP Family Conserved sequence site at 70% identity level (YKSY, AAAYCPSS, YSYFYV)

- Figures of BET V 1 allergens conserved sequence site at 70% identity level.

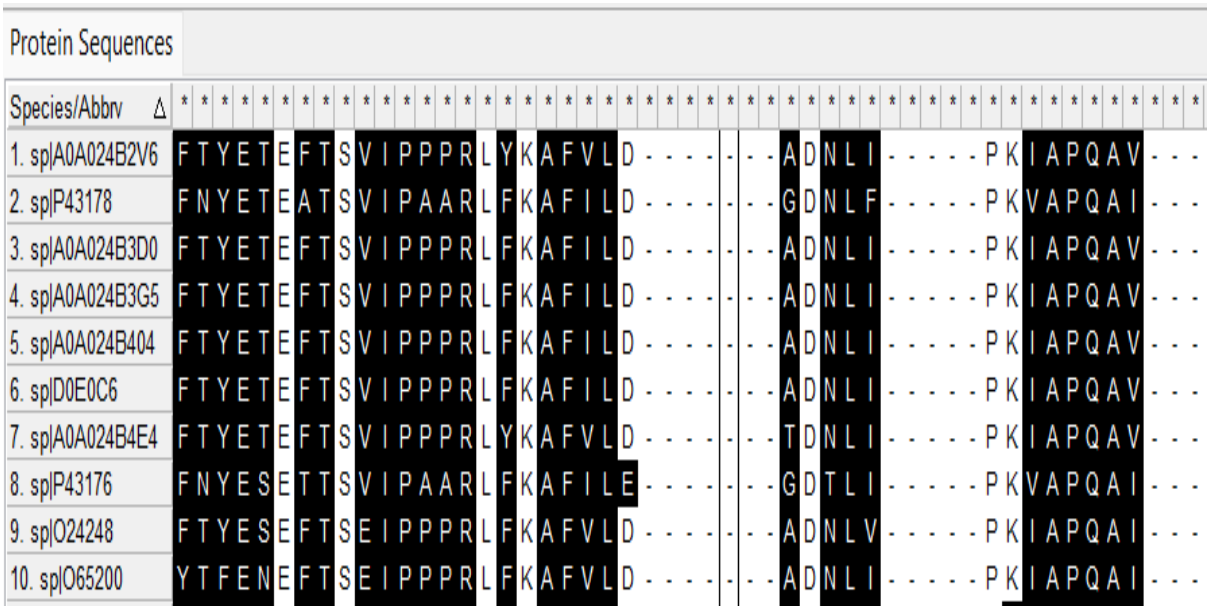


Figure 16 – BET V 1 Conserved sequence site at 70% identity level (FTYET, VIPPPR, AFIL, NLI, IAPQAV)

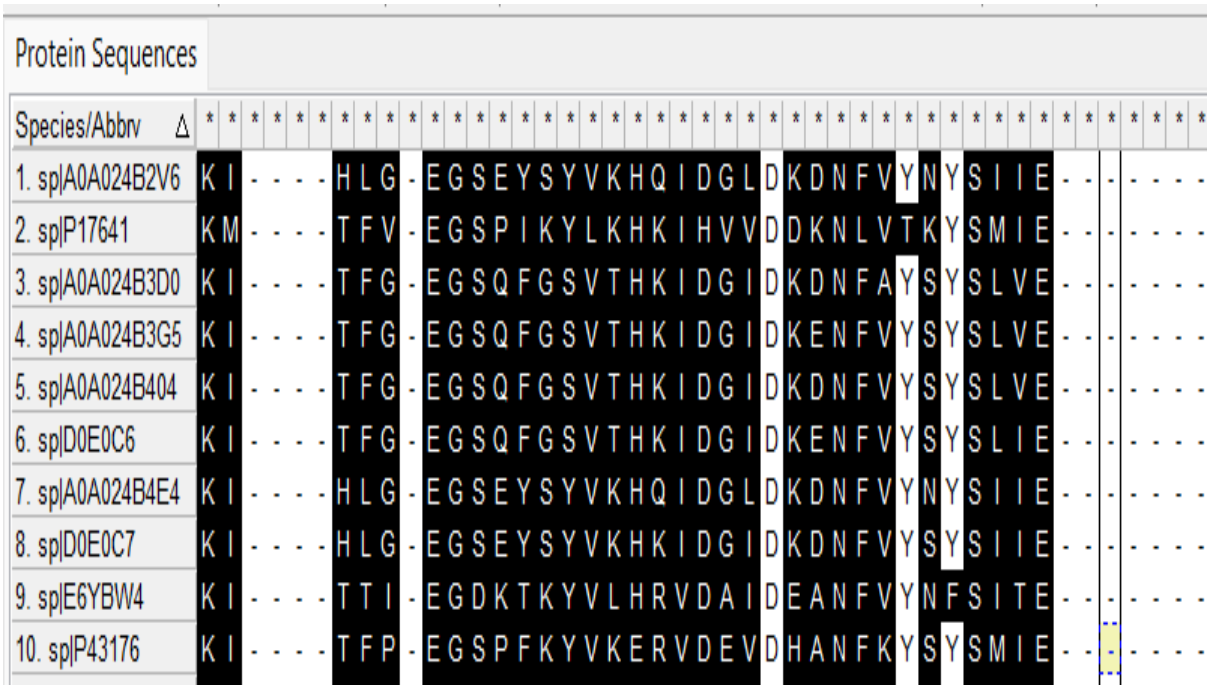


Figure 17 – BET V 1 Conserved sequence site at 70% identity level (KI, EGSQFGSVTHKIDGI)

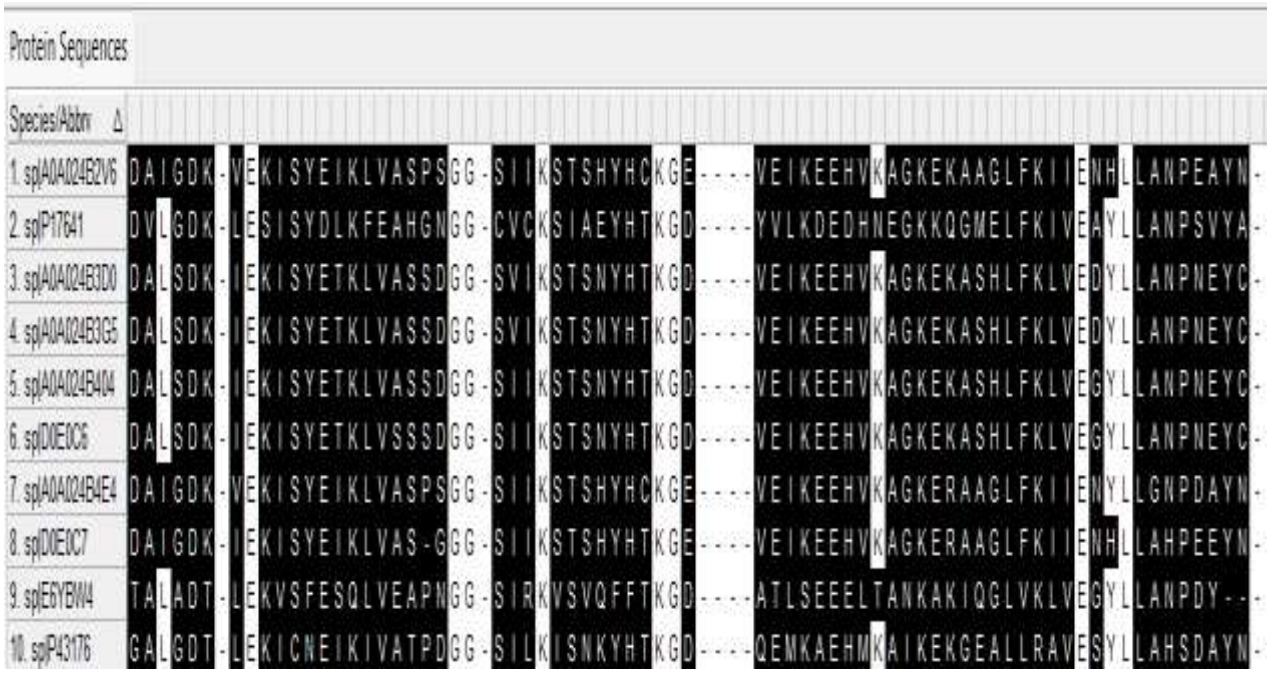


Figure 18 – BET V 1 Conserved sequence site at 70% identity level (KISYETKLVASSD, STSNYHT, VEIKEEHV, AGKEKASHLFKLV, LANPNEYC)

➢ Figures of TROPOMYOSIN allergens conserved sequence site at 70% identity level.

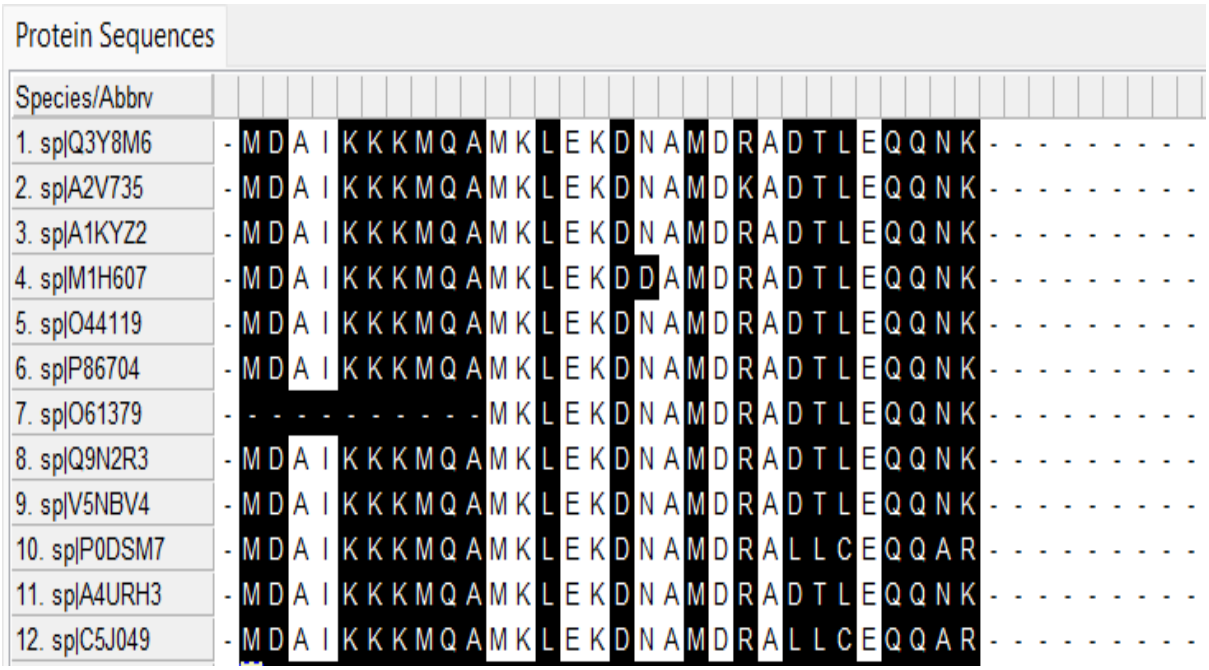


Figure 19- TROPOMYOSIN Conserved sequence site at 70% identity level (MD, KKKMQA, DTL, QQNK)

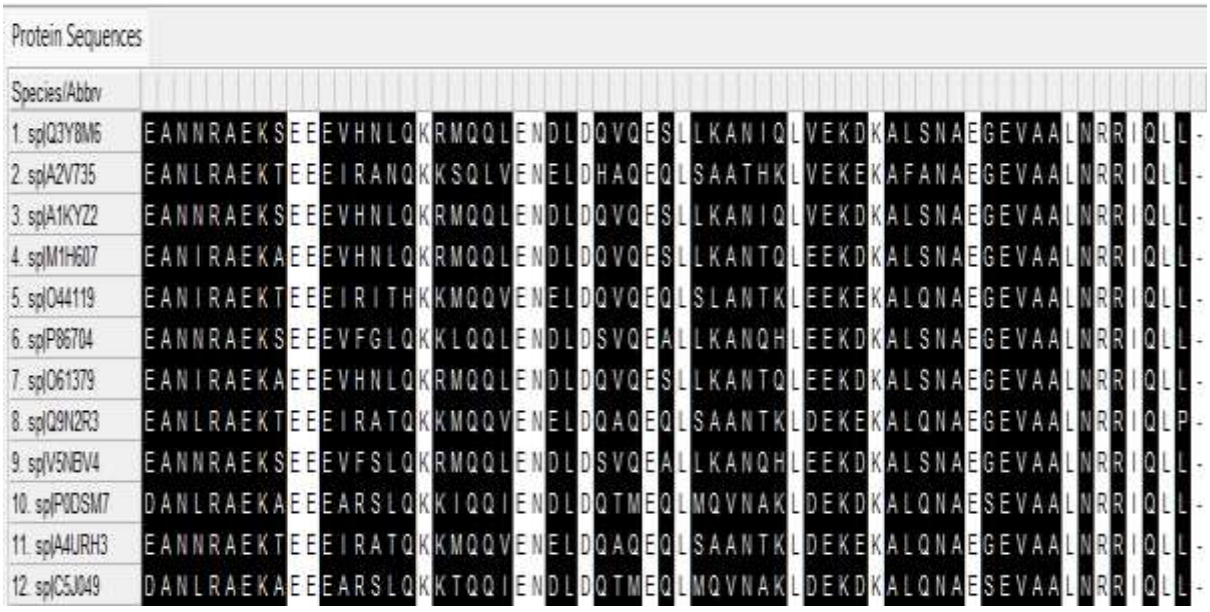


Figure 20- TROPOMYOSIN Conserved sequence site at 70% identity level (EAN, RAEK, DL, QVQ, GEVAA)

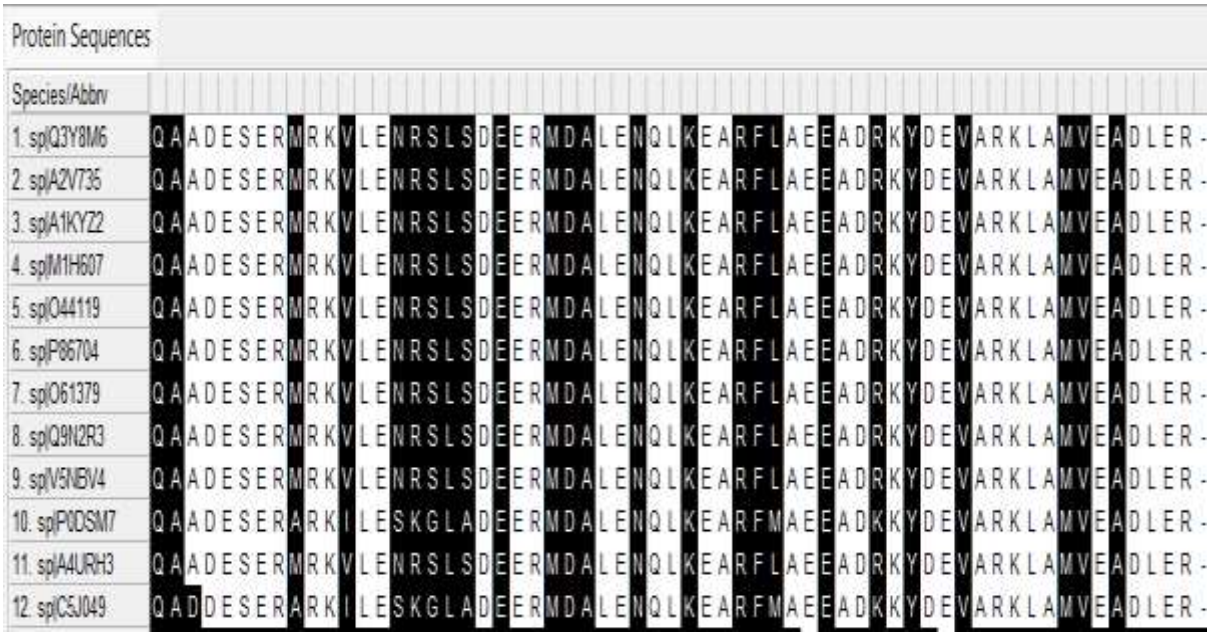


Figure 21- TROPOMYOSIN Conserved sequence site at 70% identity level (QA, NRSLS, MDA, RFL, MV)

2.2 Significance of Conserved Motifs

- The analysis of conserved motifs within each allergen family led to hypothesizing about the possible existence of functionally or allergenically critical regions.
- These motifs probably correspond to certain parts of the protein sequence that are conserved regardless of the species and are under quite significant selective pressure.

3. The Solubility of Gene Allergen Families

- The solubility comparison of allergens and non-allergens was done using the SoluProt 1.0 tool.
- The purpose of this analysis was to elucidate the characteristics of the relevant physicochemical elements that influence the solubility of allergen proteins, thereby augmenting our understanding of their functionality in various environments.

  Analysis Process

- Initially, protein sequences of both allergen and non-allergen proteins were collected from established databases.
- These sequences were then input into the SoluProt 1.0 tool for solubility prediction.
- SoluProt 1.0 utilized a sophisticated algorithm to assess solubility based on the protein's sequence.

Solubility		
	Allergen	**Non-Allergen**
Cupin	0.5264	0.7571
Bet v 1	0.8641	0.7881
Cap family	0.5158	0.6212
Tropomyosin	0.8639	0.9104

Table 5- Solubility of allergens and non-allergens of gene allergen families

- Tables of the solubility values of all allergen sequences and all non-allergen sequences of gene allergen families.

	Cupin Allergen		**Cupin Non-Allergen**	
Sr. no.	**Protein**	**Solubility**	**Protein**	**Solubility**
1	sp|A0A1L6K371	0.302	sp|E8WYN5	0.882
2	sp|A0A2I4E5L6	0.359	sp|P0C037	0.738
3	sp|A9NJG2	0.402	sp|P21816	0.947

4	sp|B3STU4	0.759	sp|P26394	0.922
5	sp|B8AL97	0.292	sp|P33487	0.275
6	sp|P04347	0.457	sp|P37745	0.595
7	sp|P04405	0.777	sp|P47096	0.657
8	sp|P04776	0.877	sp|P60334	0.947
9	sp|P11827	0.893	sp|P84140	0.838
10	sp|P25974	0.857	sp|Q03677	0.641
11	sp|Q43607	0.538	sp|Q16878	0.953
12	sp|Q8GZP6	0.618	sp|Q1LCS4	0.854
13	sp|Q8S4P9	0.743	sp|Q56185	0.939
14	sp|Q9SPL3	0.623	sp|Q562C9	0.913
15	sp|V9VGU0	0.418	sp|Q5LW89	0.432
16	sp|A0A222NNM9	0.358	sp|Q5SFD1	0.919
17	sp|B5KVH4	0.317	sp|Q6AWN0	0.786
18	sp|E3SH28	0.527	sp|Q99JT9	0.929
19	sp|F7J077	0.835	sp|Q9BV57	0.951
20	sp|P02858	0.461	sp|Q9HU21	0.915

21	sp|P0DO15	0.771	sp|Q9ZFE7	0.94
22	sp|P0DO16	0.771	sp|A0A256XLS3	0.397
23	sp|P11828	0.617	sp|A1L4T4	0.676
24	sp|Q2TPW5	0.273	sp|A3E7Z6	0.493
25	sp|Q53HY0	0.456	sp|B0JUK9	0.761
26	sp|Q6EBC1	0.483	sp|D0CY60	0.941
27	sp|Q852L2	0.288	sp|D9RZ53	0.822
28	sp|Q9SEW4	0.793	sp|O27818	0.895
29	sp|Q9XHP0	0.437	sp|O31669	0.859
30	sp|B4X640	0.516	sp|O48707	0.657
31	sp|F5B8V6	0.413	sp|O52251	0.621
32	sp|F5B8V7	0.674	sp|O86372	0.524
33	sp|F5B8V8	0.446	sp|P0A9U6	0.682
34	sp|F5B8V9	0.545	sp|P29783	0.312
35	sp|F5B8W0	0.407	sp|P32459	0.683
36	sp|F5B8W1	0.395	sp|P37780	0.758
37	sp|F5B8W2	0.364	sp|P77626	0.508
38	sp|F5B8W3	0.559	sp|P77731	0.921

39	sp|F5B8W4	0.389	sp|P83194	0.925
40	sp|F5B8W5	0.496	sp|Q0RYE0	0.913
41	sp|O23878	0.333	sp|Q10RE5	0.748
42	sp|O23880	0.375	sp|Q2RSU5	0.484
43	sp|P43238	0.838	sp|Q3ZBL1	0.884
44	sp|Q75GX9	0.231	sp|Q54ER6	0.62
45	sp|Q9SPL4	0.333	sp|Q6PBX5	0.84
46	sp|Q9XFM4	0.338	sp|Q83V26	0.664
47	sp|C0HLR7	0.517	sp|Q8GXE2	0.781
48	sp|P43237	0.799	sp|Q8H185	0.759
49			sp|Q8W108	0.805
50			sp|Q9X113	0.951
51			sp|Q9X113	0.951

Table 6- Solubility of allergens and non-allergens of CUPIN

	BET V 1 Allergen	**BET V 1 NON-Allergen**

Sr. no.	Protein	Solubility	Protein	Solubility
1	sp|Q67A25	0.77	sp|A0A2U9GGW3	0.656
2	sp|A0A1S3THR8	0.943	sp|O15155	0.487
3	sp|A4K9Z8	0.933	sp|O35152	0.338
4	sp|D0E0C7	0.953	sp|O43617	0.924
5	sp|P15494	0.954	sp|O49686	0.885
6	sp|P25816	0.928	sp|O55013	0.919
7	sp|P43185	0.958	sp|O80992	0.914
8	sp|P43211	0.885	sp|P21760	0.843
9	sp|P52778	0.932	sp|P22804	0.504
10	sp|P52779	0.874	sp|P36149	0.912
11	sp|P93330	0.927	sp|P55853	0.955
12	sp|Q256S2	0.963	sp|Q03630	0.863
13	sp|Q8H2C9	0.844	sp|Q62896	0.474
14	sp|Q9LLQ2	0.97	sp|Q84MC7	0.915
15	sp|A0A024B2V6	0.879	sp|Q8H1R0	0.931
16	sp|A0A024B3D0	0.869	sp|Q9FGM1	0.821

17	sp|A0A024B3G5	0.857	sp|Q9FJ49	0.633
18	sp|A0A024B404	0.846	sp|Q9FJ50	0.606
19	sp|A0A024B4E4	0.873	sp|Q9NRW3	0.883
20	sp|A0A161X1M2	0.803	sp|Q9NYM9	0.359
21	sp|D0E0C6	0.836	sp|Q9NZH8	0.903
22	sp|D9ZHN9	0.837	sp|Q9SN51	0.929
23	sp|E6YBW4	0.778	sp|Q9Y5R8	0.896
24	sp|G7J032	0.934	sp|A0A1S3THR8	0.943
25	sp|O04298	0.914	sp|A0A2U9GHG9	0.438
26	sp|P26987	0.958	sp|A2A1A1	0.317
27	sp|P43176	0.831	sp|A4K9Z8	0.933
28	sp|P43177	0.936	sp|D0E0C7	0.953
29	sp|P43178	0.931	sp|O04298	0.914
30	sp|P43179	0.929	sp|O35153	0.335
31	sp|P43180	0.93	sp|O35623	0.503
32	sp|P43183	0.943	sp|P0AGL5	0.493

33	sp|P43184	0.852	sp|P15494	0.954
34	sp|P43186	0.865	sp|P17446	0.269
35	sp|P45431	0.854	sp|P25816	0.928
36	sp|P80889	0.844	sp|P43185	0.958
37	sp|P85524	0.926	sp|P43211	0.885
38	sp|Q08407	0.914	sp|P52778	0.932
39	sp|Q43560	0.764	sp|P52779	0.874
40	sp|Q5ULZ4	0.944	sp|P93330	0.927
41	sp|Q7XZT8	0.923	sp|Q256S2	0.963
42	sp|Q7Y1W5	0.927	sp|Q8H2C9	0.844
43	sp|Q9AXK1	0.951	sp|Q9LLQ2	0.97
44	sp|Q9AXK2	0.962	sp|A0A024B2V6	0.879
45	sp|Q9FUW6	0.535	sp|A0A024B3D0	0.869
46	sp|Q9LLQ3	0.952	sp|A0A024B3G5	0.857
47	sp|C0HKF5	0.807	sp|A0A024B404	0.846
48	sp|O24248	0.951	sp|A0A024B4E4	0.873
49	sp|O49065	0.625	sp|A0A161X1M2	0.803

50	sp|O50001	0.885	sp|A0A3G5BB24	0.551
51	sp|O65200	0.881	sp|C0HKF5	0.807
52	sp|P13239	0.809	sp|D0E0C6	0.836
53	sp|P14710	0.827	sp|D9ZHN9	0.837
54	sp|P17641	0.74	sp|E6YBW4	0.778
55	sp|P17642	0.807	sp|G7J032	0.934
56	sp|P19417	0.776	sp|O13932	0.423
57	sp|P19418	0.721	sp|P19825	0.705
58	sp|P25985	0.901	sp|P25985	0.901
59	sp|P25986	0.9	sp|P26987	0.958
60	sp|P27047	0.845	sp|P38949	0.928
61	sp|P27538	0.762	sp|P43176	0.831
62	sp|P38948	0.827	sp|P43177	0.936
63	sp|P38949	0.928	sp|P43178	0.931
64	sp|P38950	0.933	sp|P43179	0.929
65	sp|P49372	0.943	sp|P43180	0.93
66	sp|P80890	0.829	sp|P43183	0.943

67	sp|P85126	0.887	sp|P43184	0.852
68	sp|P92918	0.805	sp|P43186	0.865
69	sp|P93105	0.782	sp|P45431	0.854
70	sp|P93333	0.771	sp|P77494	0.234
71	sp|Q05736	0.499	sp|P80889	0.844
72	sp|Q06930	0.853	sp|P80890	0.829
73	sp|Q06931	0.87	sp|P85524	0.926
74	sp|Q256S4	0.942	sp|Q08407	0.914
75	sp|Q256S6	0.934	sp|Q40280	0.892
76	sp|Q256S7	0.937	sp|Q41020	0.764
77	sp|Q40280	0.892	sp|Q43560	0.764
78	sp|Q8GT39	0.871	sp|Q5ULZ4	0.944
79	sp|Q93VR4	0.682	sp|Q7XZT8	0.923
80	sp|Q9XF37	0.831	sp|Q7Y1W5	0.927
81	sp|Q9XF38	0.866	sp|Q8FEY4	0.536
82	sp|Q9XF39	0.814	sp|Q94CG2	0.392

Table 7- Solubility of allergens and non-allergens of BET V 1

	CAP Family Allergen	**CAP Family Non-Allergen**

Sr. no.	Protein	Solubility	Protein	Solubility
1	sp|P47032	0.411	sp|A0A1I9LM04	0.551
2	sp|P47033	0.455	sp|A0A2T5Y4G4	0.558
3	sp|P54107	0.396	sp|C0VHC9	0.337
4	sp|Q9DAG6	0.244	sp|D7Y2H4	0.553
5	sp|A1BQQ5	0.319	sp|E1CG36	0.493
6	sp|P36110	0.583	sp|O15273	0.902
7	sp|P48060	0.273	sp|O24610	0.272
8	sp|P86686	0.769	sp|O43148	0.895
9	sp|Q32LB5	0.382	sp|O56129	0.464
10	sp|Q59ZX3	0.391	sp|O60573	0.876
11	sp|Q5AB48	0.56	sp|P06730	0.933
12	sp|Q6UWM5	0.258	sp|P06814	0.879
13	sp|Q7YT83	0.27	sp|P06815	0.692
14	sp|Q7ZT99	0.343	sp|P07384	0.635
15	sp|Q8JI39	0.43	sp|P0DUD5	0.649
16	sp|Q8WQ47	0.899	sp|P0DX82	0.343

17	sp|D4B327	0.614	sp|P0DX87	0.274
18	sp|D4P2Y4	0.795	sp|P12962	0.68
19	sp|E0XKJ8	0.502	sp|P15804	0.765
20	sp|F7C0L1	0.602	sp|P16259	0.872
21	sp|F8S0Y4	0.361	sp|P17213	0.318
22	sp|P08299	0.415	sp|P17555	0.916
23	sp|P09042	0.41	sp|P17655	0.844
24	sp|P0DMB9	0.753	sp|P20160	0.145
25	sp|P0DMT4	0.326	sp|P20482	0.411
26	sp|P0DPU0	0.337	sp|P20807	0.886
27	sp|P0DPU1	0.266	sp|P32783	0.531
28	sp|P0DPU2	0.437	sp|P35750	0.603
29	sp|P0DPU5	0.355	sp|P36621	0.557
30	sp|P0DPV2	0.322	sp|P40121	0.918
31	sp|P10736	0.459	sp|P40123	0.772
32	sp|P10737	0.717	sp|P40124	0.743

33	sp|P35759	0.606	sp|P52298	0.919
34	sp|P35778	0.5	sp|P54654	0.897
35	sp|P35783	0.917	sp|P63073	0.932
36	sp|P35793	0.384	sp|P80015	0.106
37	sp|P60623	0.611	sp|P83774	0.733
38	sp|P79845	0.399	sp|P83776	0.592
39	sp|P86870	0.491	sp|P83782	0.756
40	sp|Q00008	0.218	sp|Q01518	0.645
41	sp|Q05110	0.823	sp|Q03380	0.332
42	sp|Q09GJ9	0.465	sp|Q07009	0.903
43	sp|Q52PV9	0.911	sp|Q09161	0.91
44	sp|Q59PV6	0.318	sp|Q12052	0.376
45	sp|Q5AB49	0.474	sp|Q13541	0.904
46	sp|Q7Z156	0.803	sp|Q53F19	0.699
47	sp|Q7ZTA0	0.334	sp|Q6PKG0	0.804
48	sp|Q7ZZN9	0.457	sp|Q8BZR9	0.671
49	sp|Q8JI40	0.503	sp|Q8TEQ6	0.394

50	sp|Q93YG7	0.78	sp|Q9D0L8	0.742
51	sp|Q962V9	0.899	sp|Q9KVG5	0.52
52	sp|Q9CQ35	0.732	sp|Q9KVG6	0.242
53	sp|A0A218QX58	0.647	sp|Q9V3L6	0.716
54	sp|A9QQ26	0.25	sp|Q9XFD1	0.636
55	sp|A9YME1	0.687	sp|A0A2K8K5C5	0.39
56	sp|B0VXV6	0.385	sp|A0A381HAP5	0.524
57	sp|B2MVK7	0.42	sp|A0A5D0EMF2	0.527
58	sp|B7FDI0	0.509	sp|P0DTF2	0.355
59	sp|B7FDI1	0.498	sp|P0DTF3	0.512
60	sp|B9URJ1	0.353	sp|P0DTF4	0.594
61	sp|C0ITL3	0.55	sp|P0DUD9	0.47
62	sp|P0DSI3	0.401	sp|P0DUE0	0.554
63	sp|P11670	0.342	sp|P0DXA4	0.288
64	sp|P35760	0.93	sp|P19198	0.474
65	sp|P35779	0.668	sp|P29195	0.778
66	sp|P35780	0.772	sp|P52481	0.838

67	sp|P35781	0.79	sp|P94017	0.597
68	sp|P35782	0.811	sp|Q08163	0.745
69	sp|P35784	0.929	sp|Q8GX47	0.717
70	sp|P35785	0.926	sp|Q8S9J8	0.674
71	sp|P35786	0.824	sp|Q9CYT6|	0.87
72	sp|P35787	0.874	sp|Q9SA65|	0.718
73	sp|P35792	0.368	sp|P40122|	0.561
74	sp|P35794	0.353	sp|Q3SYV4	0.711
75	sp|P35795	0.4		
76	sp|P81656	0.629		
77	sp|P81657	0.764		
78	sp|P83377	0.786		
79	sp|P85840	0.696		
80	sp|P85860	0.468		
81	sp|Q05108	0.765		
82	sp|Q05109	0.768		
83	sp|Q05968	0.36		
84	sp|Q08697	0.426		
85	sp|Q16937	0.573		

86	sp|Q2L6Z1	0.73		
87	sp|Q2XXP1	0.471		
88	sp|Q2XXP2	0.472		
89	sp|Q2XXQ0	0.435		
90	sp|Q2XXQ8	0.259		
91	sp|Q2XXR1	0.272		
92	sp|Q2XXR2	0.227		
93	sp|Q40374	0.288		
94	sp|Q59WG5	0.302		
95	sp|Q93YI9	0.808		
96	sp|Q9CWG1	0.313		
97	sp|Q9XSD3	0.199		
98	sp|W4VS53	0.61		

Table 8- Solubility of allergens and non-allergens of CAP Family

	TROPOMYOSIN Allergen		**TROPOMYOSIN Non-Allergen**	
Sr. no.	**Protein**	**Solubility**	**Protein**	**Solubility**
1	sp|Q3Y8M6	0.878	sp|P04268|	0.965

2	sp|A2V735	0.857	sp|P04692|	0.962
3	sp|C7E3T4	0.882	sp|P06753|	0.924
4	sp|Q8WQ47	0.899	sp|P06754|	0.899
5	sp|A1KYZ2	0.878	sp|P07951|	0.959
6	sp|M1H607	0.881	sp|P09493|	0.958
7	sp|O44119	0.874	sp|P09495|	0.867
8	sp|P86704	0.887	sp|P21107|	0.927
9	sp|Q2V0V2	0.806	sp|P58771|	0.962
10	sp|Q9N2R3	0.824	sp|P58774|	0.96
11	sp|V5NBV4	0.885	sp|P58775|	0.96
12	sp|A4URH3	0.861	sp|P67936|	0.871
13	sp|A7UMC0	0.908	sp|Q22866|	0.884
14	sp|C5J049	0.883	sp|Q27249|	0.857
15	sp|D2DGW3	0.896	sp|Q3Y8M6|	0.878
16	sp|O18416	0.832	sp|Q63610|	0.868
17	sp|O61379	0.886	sp|Q6IRU2|	0.87
18	sp|O96764	0.861	sp|A1KYZ2|	0.878

19	sp|O97192	0.838	sp|A2V735|	0.857
20	sp|P0DSM6	0.886	sp|P02561|	0.873
21	sp|P0DSM7	0.886	sp|P09491|	0.82
22	sp|Q1A7B1	0.827	sp|P19352|	0.966
23	sp|Q1A7B2	0.713	sp|P42639|	0.945
24	sp|Q1A7B3	0.723	sp|P58772|	0.962
25	sp|Q23939	0.844	sp|P67937|	0.871
26	sp|Q25456	0.888	sp|P84335|	0.941
27	sp|Q3BJY7	0.838	sp|Q2V0V2|	0.806
28	sp|Q7M3Y8	0.952	sp|Q5KR49|	0.959
29	sp|Q8T380	0.902	sp|O02389|	0.882
30	sp|Q95WY0	0.887	sp|P13104|	0.954
31	sp|Q9GZ69	0.901	sp|P58773|	0.965
32	sp|Q9GZ70	0.903	sp|Q26503|	0.867
33	sp|Q9GZ71	0.834	sp|Q26519|	0.868
34	sp|Q9NAS5	0.896	sp|Q5KR47|	0.921
35	sp|Q9NFZ4	0.843	sp|Q5KR48|	0.96

Table 9- Solubility of allergens and non-allergens of TROPOMYOSIN

- From the solubility analysis of allergens and non-allergens, the average solubility of Cupin allergens was (0.5264), while non-allergens had an average solubility of (0.7571). For Bet v 1 allergens, the average solubility was (0.8641), compared to (0.7881) for non-allergens. The solubility average of Cap family allergens was (0.5158), whereas non-allergens had an average of (0.6212). The solubility of tropomyosin allergens was (0.8639), and for non-allergens, it had been (0.9104).
- In the analysis, 48 allergen sequences and 51 non-allergen sequences were used for Cupin, 82 allergen sequences and 82 non-allergen sequences for Bet v 1, 98 allergen sequences and 74 non-allergen sequences for the Cap family, and 35 allergen sequences and 35 non-allergen sequences for tropomyosin.

CHAPTER 5
DISCUSSION

From the two-dimensional CoAn biplot graphs (Figures 1, 2, 3, and 4) that are shown in the following figures, noticeable separations between allergens and non-allergens were observed with regard to all the studied allergen families, inclusive of cupin, Bet v 1, Cap family, and tropomyosin.As a result, it was determined that there were confirmed differences in the sequences of allergens and non-allergens. As seen in the case of the biplot, there is segregation between the amino acid compositions and sequence aspects of allergens and non-allergens in each allergen group. Based on this discovery, it is evident that the CoAn biplot analysis is effective in revealing structural and compositional differences between allergenic and non-allergenic proteins and can aid in future classification and strategies for managing allergens. Since these all point to future directions, future research could expand on these distinctions to

improve allergen identification and allergen avoidance in various food and environmental settings.

There were differences in the use of amino acids between allergen and non-allergen sequences that were statistically significant ($p < 0.05$) in a number of protein families. In cupin proteins, allergens typically showed more underrepresented than overrepresented amino acids, suggesting regulatory roles in allergenicity. Conversely, CAP family allergens consistently exhibited overrepresented amino acids, indicating potential structural or functional significance. However, meaningful comparisons between Bet v 1 and tropomyosin were hindered due to the absence of cysteine in non-allergenic Bet v 1 sequences and allergenic tropomyosin sequences, which are crucial for protein stability and structural integrity. These findings underscore the varied amino acid compositions influencing allergenic potential across different protein families.

The conserved motifs found at the 70% identity level give essential information about the structural and functional characteristics of allergens from the various families. In cupin, Bet v 1, CAP (cysteine-rich secretory), and tropomyosin allergens, there are 18, 12, 18, and 14 conserved motifs in total. It will be quite beneficial to make the assumption that these motifs have significant functions with reference to protein stability, allergenicity, and interaction with the immune system. Knowledge of these conserved motifs provides imprints for allergen-specific therapies and diagnostic approaches.

Allergens in the cupin, CAP, and tropomyosin families exhibited lower solubility than non-allergenic counterparts, while allergens from the Bet v 1 family showed higher solubility (see Table 5). These solubility differences likely influence allergen stability and interaction with the immune system, highlighting their relevance in understanding allergenic properties across protein families.

CHAPTER 6

CONCLUSION

Based on the findings of the study, several important conclusions can be drawn. The RAAU data correspondence plot also shows how the allergen and non-allergen families use amino acids in different ways across all allergen groups. This clearly indicates that both allergen and non-allergen families consist of distinct components, consistently exhibiting differences. The developed visualization clearly shows that allergens and non-allergens occupy separate areas in the correspondence analysis space, indicating significant differences in amino acid preferences and distribution before and after correspondence analysis. These differences may be related to allergens' biological activities and allergenicity.

It was also observed that all the allergens have conserved motifs that could point to a possible molecular pattern that may be associated with allergenicity. These are amino acid sections that do not differ much across an allergen group's proteins, suggesting functional domains; they

may be critical structural components or immune-related. The identification of basic motifs in a broad spectrum of allergens reveals a specific feature of these proteins, which may be considered in further allergenicity studies.

The results of the solubility analysis indicate that some allergens have high solubility while others of the same family have low solubility. In particular, allergens from the Cupin, CAP, and tropomyosin families are less soluble than non-allergenic homologs. It is noteworthy that the allergens belonging to the Bet v 1 groups exhibit higher solubility in comparison to other proteins that are not allergenic. Such solubility profiles are well aligned with the structural differences that make an allergenic protein allergenic across different allergen families. The lower solubility of Cupin, CAP, and tropomyosin allergens may be linked to their ability to form stable aggregates, which enhances their resistance to degradation and increases their allergenic potential. However, higher solubility in Bet v 1 allergens could enhance their propensity to disperse in the body and with the immune system, hence their allergenic effect.

These conclusions provide a comprehensive understanding of the molecular characteristics that differentiate allergens from non-allergens, emphasizing the role of amino acid usage patterns, conserved motifs, and solubility profiles in influencing allergenic properties. The findings contribute valuable insights for future research aimed at elucidating the mechanisms of allergenicity and developing strategies for allergy prevention and treatment.

CHAPTER 7

REFERENCES

1. Singh, B., Ahanathapillai, V., Sharma, N. R., Jan, S., Roy, A., & Upadhyay, A. K. (2022). Structural insights into the amino acid usage variations in the profilin gene family. *Amino Acids*, *54*(3), 411–419. https://doi.org/10.1007/s00726-022-03138-4
2. Anvari, S., Miller, J., Yeh, C. Y., & Davis, C. M. (2019). IgE-Mediated Food Allergy. *Clinical Reviews in Allergy and Immunology*, *57*(2), 244–260. https://doi.org/10.1007/s12016-018-8710-3
3. Rojo-Gutiérrez, M. I., Ballesteros-González, D., & Ortiz-Durán, A. K. (2023, October 1). Non-IgE-mediated food allergy. *Revista Alergia Mexico*. Colegio Mexicano de Inmunologia Clinica y Alergia A.C.
4. González-Díaz, S. N. (2023, October 1). Epidemiology of food allergy. *Revista Alergia Mexico*. Colegio Mexicano de Inmunologia Clinica y Alergia A.C. https://doi.org/10.29262/ram.v70i4.1328
5. Izabela, K. M., Emilia, M., Weronika, G., Danuta, K., Jakub, S., Beata, B. J., … Bożena, C. (2024). The Sensitization Profile for Selected Food Allergens in Polish Children

Assessed with the Use of a Precision Allergy Molecular Diagnostic Technique. *International Journal of Molecular Sciences*, *25*(2). https://doi.org/10.3390/ijms25020825

6. Kobelkova, I. V., Korosteleva, M. M., Nikityuk, D. B., & Kobelkova, M. S. (2021). GLYCATION END PRODUCTS AND TECHNOLOGICAL ASPECTS OF REDUCING IMMUNOGENICITY OF SPECIALIZED FOOD PRODUCTS FOR NUTRITION OF ATHLETES. *Pisevye Sistemy/Food Systems*, *4*(4), 278–285. https://doi.org/10.21323/2618-9771-2021-4-4-278-285
7. Tedner, S. G., Asarnoj, A., Thulin, H., Westman, M., Konradsen, J. R., & Nilsson, C. (2022, March 1). Food allergy and hypersensitivity reactions in children and adults—A review. *Journal of Internal Medicine*. John Wiley and Sons Inc. https://doi.org/10.1111/joim.13422
8. Loh, W., & Tang, M. L. K. (2018). The epidemiology of food allergy in the global context. *International Journal of Environmental Research and Public Health*, *15*(9). https://doi.org/10.3390/ijerph15092043
9. Scarpone, R., Kimkool, P., Ierodiakonou, D., Leonardi-Bee, J., Garcia-Larsen, V., Perkin, M. R., & Boyle, R. J. (2023). Timing of Allergenic Food Introduction and Risk of Immunoglobulin E-Mediated Food Allergy: A Systematic Review and Meta-analysis. *JAMA Pediatrics*, *177*(5), 489–497. https://doi.org/10.1001/jamapediatrics.2023.0142
10. van der Valk, J. P. M., Schreurs, M. W. J., el Bouch, R., Arends, N. J. T., & de Jong, N. W. (2016, September 1). Mono-sensitisation to peanut component Ara h 6: a case series of five children and literature review. *European Journal of Pediatrics*. Springer Verlag. https://doi.org/10.1007/s00431-016-2733-7
11. Wang, X., Ding, J., Chen, L., Wang, H., & Wang, X. (2024). Study on allergen components of birch pollen in Beijing area. *Lin Chuang Er Bi Yan Hou Tou Jing Wai Ke Za Zhi = Journal of Clinical Otorhinolaryngology, Head, and Neck Surgery*, *38*(3), 195–199. https://doi.org/10.13201/j.issn.2096-7993.2024.03.003
12. Smolnikov, E. V., Litovkina, A. O., & Byazrova, M. G. (2023). Development of cofactor-dependent anaphylaxis in a patient with cross-reactive food allergy. *Russian Journal of Allergy*, *20*(4), 543–548. https://doi.org/10.36691/RJA16898
13. Kamath, S. D., Bublin, M., Kitamura, K., Matsui, T., Ito, K., & Lopata, A. L. (2023, May 1). Cross-reactive epitopes and their role in food allergy. *Journal of Allergy and Clinical Immunology*. Elsevier Inc. https://doi.org/10.1016/j.jaci.2022.12.827

14. McKendry, R. T., Kwok, M., Hemmings, O., James, L. K., & Santos, A. F. (2021). Allergen-specific IgG show distinct patterns in persistent and transient food allergy. *Pediatric Allergy and Immunology*, *32*(7), 1508–1518. https://doi.org/10.1111/pai.13567
15. Arasi, S., Nurmatov, U., Dunn-Galvin, A., Roberts, G., Turner, P. J., Shinder, S. B., … Fiocchi, A. (2023). WAO consensus on DEfinition of Food Allergy SEverity (DEFASE). *World Allergy Organization Journal*, *16*(3). https://doi.org/10.1016/j.waojou.2023.100753
16. Francis, O. L., Wang, K. Y., Kim, E. H., & Moran, T. P. (2020). Common food allergens and cross-reactivity. *Journal of Food Allergy*, *2*(1), 17–21. https://doi.org/10.2500/jfa.2020.2.200020
17. Kazatsky, A. M., & Wood, R. A. (2016, March 1). Classification of Food Allergens and Cross-Reactivity. *Current Allergy and Asthma Reports*. Current Medicine Group LLC 1. https://doi.org/10.1007/s11882-016-0601-1
18. Nuyttens, L., De Vlieger, L., Diels, M., Schrijvers, R., & Bullens, D. M. A. (2023). The clinical and immunological basis of early food introduction in food allergy prevention. *Frontiers in Allergy*. Frontiers Media S.A. https://doi.org/10.3389/falgy.2023.1111687
19. Peters, R. L., Krawiec, M., Koplin, J. J., & Santos, A. F. (2021, May 1). Update on food allergy. *Pediatric Allergy and Immunology*. Blackwell Publishing Ltd. https://doi.org/10.1111/pai.13443
20. Wang, J., Yu, Y., Zhao, Y., Zhang, D., & Li, J. (2013). Evaluation and integration of existing methods for computational prediction of allergens. *BMC Bioinformatics*, *14*(SUPPL4). https://doi.org/10.1186/1471-2105-14-S4-S1
21. Dimitrov, I., Flower, D. R., & Doytchinova, I. (2013). AllerTOP - a server for in silico prediction of allergens. *BMC Bioinformatics*, *14*(SUPPL6). https://doi.org/10.1186/1471-2105-14-S6-S4
22. Denery-Papini, S., Lollier, V., Mameri, H., Pietri, M., Larre, C., Gaudin, J.-C., … Tessier, D. (2014). In silico prediction of B cell epitopes and experimental validation on wheat allergens. *Clinical and Translational Allergy*, *4*(S2). https://doi.org/10.1186/2045-7022-4-s2-o4
23. Jiang, B., Qu, H., Hu, Y., Ni, T., & Lin, Z. (2007). Computational analysis of the relationship between allergenicity and digestibility of allergenic proteins in simulated gastric fluid. *BMC Bioinformatics, 8*. https://doi.org/10.1186/1471-2105-8-375

24. Musil, M., Konegger, H., Hon, J., Bednar, D., Damborsky, J. (2019). Computational Design of Stable and Soluble Biocatalysts. *ACS Catalysis* 9: 1033−1054.
25. Berman, H. M., Gabanyi, M. J., Kouranov, A., Micallef, D. I., Westbrook, J., Protein Structure Initiative Network Of Investigators (2017). Protein Structure Initiative – Targettrack 2000–2017.
26. Price, W. N., Handelman, S. K., Everett, J. K., Tong, S. N., Bracic, A., Luff, J. D., Hunt, J. F. (2011). Large-Scale Experimental Studies Show Unexpected Amino Acid Effects on Protein Expression and Solubility *in vivo* in *E. coli*. *Microbial Informatics and Experimentation* 1: 6.
27. Lim, S. J., Tong, J. C., Chew, F. T., & Tammi, M. T. (2008). The value of position-specific scoring matrices for assessment of protein allegenicity. *BMC Bioinformatics*, *9*(S12). https://doi.org/10.1186/1471-2105-9-s12-s21
28. Fiers, M. W. E. J., Kleter, G. A., Nijland, H., Peijnenburg, A. A. C. M., Nap, J. P., & van Ham, R. C. H. J. (2004). Allermatch™, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics*, *5*. https://doi.org/10.1186/1471-2105-5-133
29. de la Cruz, S., López-Calleja, I., Martín, R., González, I., Alcocer, M., & García, T. (2017). Recent advances in the detection of allergens in foods. In *Methods in Molecular Biology* (Vol. 1592, pp. 263–295). Humana Press Inc. https://doi.org/10.1007/978-1-4939-6925-8_20
30. Radauer, C., et al. "Molecular and immunological characterization of novel weed pollen pan-allergens." *Allergy*, vol. 66, no. 11, 2011, pp. 1429-1437.
31. Valenta, R., Hochwallner, H., Linhart, B., & Pahr, S. (2015). Food allergies: The basics. *Gastroenterology*, *148*(6), 1120-1131.e4. https://doi.org/10.1053/j.gastro.2015.02.006
32. Bublin, M., Eiwegger, T., & Breiteneder, H. (2014). Do lipids influence the allergic sensitization process? *Journal of Allergy and Clinical Immunology*, *134*(3), 521–529. https://doi.org/10.1016/j.jaci.2014.04.015
33. Mastrorilli, C., Caffarelli, C., & Hoffmann-Sommergruber, K. (2017, December 1). Food allergy and atopic dermatitis: Prediction, progression, and prevention. *Pediatric Allergy and Immunology*. Blackwell Publishing Ltd. https://doi.org/10.1111/pai.12831
34. Tosca, M. A., Schiavetti, I., Olcese, R., Trincianti, C., & Ciprandi, G. (2023). Molecular Allergy Diagnostics in Children with Cow's Milk Allergy: Prediction of Oral Food

Challenge Response in Clinical Practice. *Journal of Immunology Research, 2023*. https://doi.org/10.1155/2023/1129449

35. Costa, J., Villa, C., Verhoeckx, K., Cirkovic-Velickovic, T., Schrama, D., Roncada, P., … Holzhauser, T. (2022, February 1). Are Physicochemical Properties Shaping the Allergenic Potency of Animal Allergens? *Clinical Reviews in Allergy and Immunology*. Springer. https://doi.org/10.1007/s12016-020-08826-1
36. Sicherer, S. H., & Sampson, H. A. (2014). Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *Journal of Allergy and Clinical Immunology*, 133(2), 291-307. https://doi.org/10.1016/j.jaci.2013.11.020

DrillBit

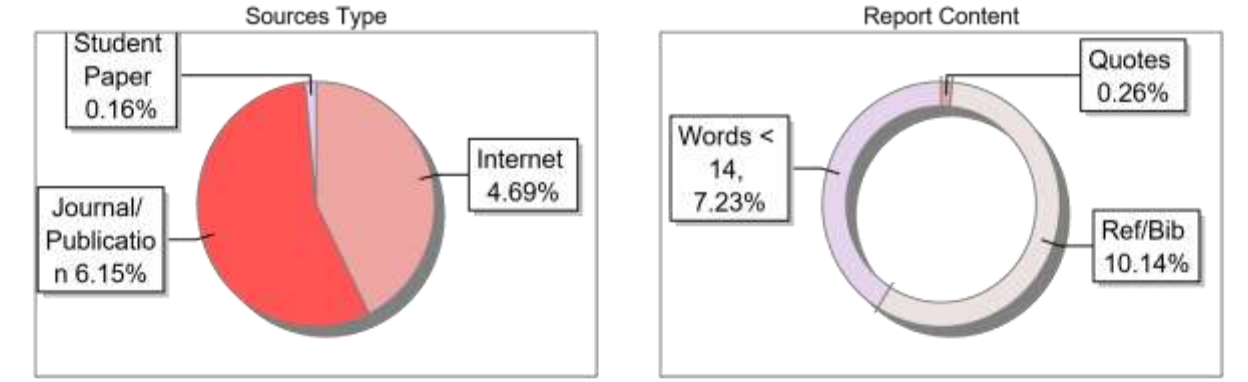
The Report is Generated by DrillBit Plagiarism Detection Software

Submission Information

Author Name	Devrat
Title	MSc thesis
Paper/Submission ID	2162418
Submitted by	atul.upadhyay@thapar.edu
Submission Date	2024-07-25 10:38:24
Total Pages, Total Words	79, 11523
Document type	Thesis

Result Information

Similarity **11 %**



Exclude Information

Quotes	Not Excluded	Language	English
References/Bibliography	Not Excluded	Student Papers	Yes
Source: Excluded < 14 Words	Not Excluded	Journals & publishers	Yes
Excluded Source	**0 %**	Internet or Web	Yes
Excluded Phrases	Not Excluded	Institution Repository	Yes

Database Selection

A Unique QR Code use to View/Download/Share Pdf File