Allergologia et immunopathologia
Sociedad Española de Inmunología Clínica, Alergología y Asma Pediátrica
www.all-imm.com

ORIGINAL ARTICLE

OPEN ACCESS

Using the A-RISC index to predict IgE cross-reactivity in legume allergens: An in silico approach

Jackelyn Stephanny Páez Velásquez[a,b]*

[a]*Pediatric Allergist and Immunologist, Fundación Neumológica Colombiana, Bogotá, Colombia*
[b]*PhD Student in Biomedical Basic Sciences, Universidad El Bosque, Bogotá, Colombia*

Received 30 June 2025; Accepted 3 August 2025
Available online 20 October 2025

KEYWORDS
Legume Allergens;
Cross-reactivity;
A-RISC Index;
In Silico;
Component-resolved Diagnosis;
Food Allergy

Abstract
Introduction: Legumes are a common source of allergic sensitization in many regions worldwide. Structural similarity among homologous proteins can lead to IgE-mediated cross-reactivity. In this context, in silico analysis offers a valuable approach to predict potential molecular interactions among related allergens and to support the interpretation of risk in patients with multiple sensitizations.
Methods: An in silico analysis was conducted to evaluate sequence homology, structural conservation, and surface exposure of IgE epitopes across five major protein families: 11S globulins, 7S globulins, 2S albumins, nsLTPs, and PR-10. Tools included multiple sequence alignment, A-RISC index calculation, and 3D visualization with ChimeraX.
Results: PR-10 proteins exhibited high homology (A-RISC >0.75), suggesting a high risk of cross-reactivity. Vicilins and glycinins showed intermediate similarity (A-RISC 0.45–0.57), while nsLTPs and 2S albumins displayed low A-RISC values (<0.50), although conserved structural motifs were identified in immunologically relevant regions.
Conclusion: This in silico approach enables early identification of cross-reactivity potential, reinforces the value of component-resolved molecular diagnostics, and contributes to improved food labeling, clinical decision-making, and nutritional safety in patients with multiple sensitizations.

***Corresponding author**: Jackelyn Stephanny Páez Velásquez, Fundación Neumológica Colombiana, Bogotá, Colombia; Biomedical Basic Sciences, Universidad El Bosque, Bogotá, Colombia. *Email address:* drapaezaalergo@gmail.com

https://doi.org/10.15586/aei.v53sp2.1455

Glossary

- **Allergen:** A molecular component with immunogenic potential. Derived from natural kingdoms, they belong to specific allergenic protein families based on their amino acid sequence and structural similarity.
- **Allergy:** A clinical manifestation triggered by contact with an allergen capable of inducing an inflammatory immune response.
- **A-RISC (Allergens' Relative Identity, Similarity, and Cross-reactivity):** An index calculated based on homology, similarity, and identity between two or more allergens. It estimates the probability of cross-reactivity among members of the same family.
- **Epitope**: The antigenic determinant of a macromolecule recognized by specific immunoglobulins (e.g., IgE).
- **Homology:** The evolutionary relationship between two sequences. Homologous proteins are commonly recognized by specific IgE. Homology is calculated using the A-RISC index.
- **Identity:** The degree of exact amino acid match between two different protein sequences.
- **Similarity:** The percentage of residues with similar physicochemical properties (size, charge, hydrophobicity).
- **Cross-reactivity:** An immune response induced by an allergen different from the primary sensitizer. Sequence identity ≥70% is commonly indicative of potential cross-reactivity.

Introduction

Food allergies represent a growing public health concern, affecting a significant proportion of the global population, with an estimated prevalence of 6–8% in children and 3–4% in adults, depending on age and geographic region.[1] Within this group, legumes are a frequent cause of allergic sensitization, with soy and peanut being the primary culprits.[2] In Latin America, an increasing prevalence of food allergies has been reported, with higher sensitization rates to legumes such as lentils, beans, and chickpeas, especially in pediatric populations.[3]

Besides being a vital source of plant-based proteins, legumes contain allergenic proteins with high structural similarity, which promotes cross-reactivity among species.[4] These proteins are classified into various superfamilies, including 7S and 11S globulins, 2S albumins, oleosins, and pathogenesis-related proteins (PR-10), which are known for their stability and digestive resistance, characteristics that contribute to their clinical allergenic potential.[5]

Accurate food labeling plays a critical role in preventing adverse reactions among individuals with food allergies. International regulatory agencies, such as the European Food Safety Authority (EFSA) and the US Food and Drug Administration (FDA), mandate the explicit declaration of allergenic ingredients in processed food products.[6] However, the implementation and enforcement of these labeling standards remain inconsistent, particularly in plant-derived foods and within developing countries, where regulatory oversight and infrastructure may be limited.

Molecular diagnostic strategies have become increasingly important in food allergy characterization. Tools such as the *Molecular Allergology User's Guide* (MAUG 2.0)[7] have improved the identification of specific proteins using microarrays and peptide sequencing, enabling the differentiation between primary sensitization and cross-reactivity.

Considering the importance of legume-related allergies and the need to enhance both diagnostic and labeling strategies, this study employs in silico tools to analyze the sequence identity of allergenic proteins in legumes. The results are expected to contribute to a deeper understanding of cross-reactivity and provide critical insights for clinical management of allergic patients, as well as inform preventive strategies in the food industry.[8]

The A-RISC index is a valuable method for estimating the probability of cross-reactivity between two allergens belonging to the same protein family.

A similarity matrix was constructed using the **A-RISC (Allergens'- Relative Identity, Similarity, and Cross-reactivity)** index, which reflects structural homology between pairs of proteins. The score ranges from 0 to 1 and is interpreted as follows:

- **0.70–1.00:** *High risk of IgE cross-reactivity*
- **0.50–0.69:** *Moderate risk of IgE cross-reactivity*
- **< 0.50:** *Low risk of IgE cross-reactivity.*

Higher scores indicate an increased likelihood of immunological recognition between allergens, with implications for component-resolved diagnosis (CDR) and immunotherapy strategies.

According to the index developed by Chruszcz et al. (2018), the **A-RISC** combines amino acid identity and

Table 1 Heatmap of A-RISC index among legume allergenic proteins.

Heatmap	Score ranges	Interpretation
0-0.09 0.1-0.19 0.2-0.29 0.30-0.39 0.4-0.49	<0.50	**Low risk** of IgE cross-reactivity
0.5-0.59 0.6-0.69 0.7-0.79	0.50–0.74	**Moderate risk** of IgE cross-reactivity
0.8-0.89 0.9-1	≥0.75	**High risk** of IgE cross-reactivity

A-RISC[9]

similarity to estimate the probability that an IgE antibody will fail to distinguish between two allergens from the same protein family. It is based on a weighted average of:

- **% Identity:** The proportion of exact amino acid matches at aligned positions.
- **% Similarity:** The proportion of chemically similar amino acids, calculated using the BLOSUM62 substitution matrix.

The resulting **A-RISC score ranges from 0 to 1**, where:

- **0.00** indicates no structural relationship.
- **1.00** corresponds to identical sequences.

Methods

Allergenic protein sequences were retrieved from specialized databases including the **WHO/IUIS Allergen Nomenclature Home Page**, **AllFam**, **NCBI**, and **UniProt**. Only full-length, mature protein sequences were considered; signal peptides and propeptides were excluded. For officially recognized allergens, standardized nomenclature was used, including four digits after the period (e.g., Ara h 2.0101). In specific cases where the crystallizable structure of the primary isoform was unavailable, closely related isoforms (e.g., 0.0101) were accepted.

The main allergens analyzed in this study included:

- Soybean (*Glycine max*)
- Lupin (*Lupinus spp.*)
- Pea (*Pisum sativum*)
- Peanut (*Arachis hypogaea*)
- Lentil (*Lens culinaris*)
- Chickpea (*Cicer arietinum*)
- Common bean (*Phaseolus vulgaris*)

For *the* in silico analysis, multiple sequence alignments were performed using **Clustal Omega** (https://www.ebi.ac.uk/jdispatcher/msa) to identify significant alignment regions across allergenic proteins. Subsequently, **SIAS** (http://imed.med.ucm.es/Tools/sias.html) was employed with the Blosum62 substitution matrix to calculate pairwise identity and similarity percentages. The resulting similarity matrices were comparable to those used in previous allergen cross-reactivity studies involving aeroallergens, fruits, nuts, shellfish, and fish.

To assess structural homology and potential cross-reactivity risk, **A-RISC** indices were computed, and corresponding heatmaps were generated for visual representation.[9,10]

Additionally, family-specific multiple sequence alignments were visualized using **Jalview**, applying Blosum62-based coloring and a 90% conservation threshold, to highlight structurally conserved regions or domains.

Results

Protein family classification

Only protein families with at least three allergenic sequences derived from different food sources were included in the analysis.

Families such as profilins, oleosins, defensins, and cyclophilins were excluded because representative allergens were identified in only one or two plant species, limiting the possibility of performing a meaningful comparative analysis of cross-reactivity. The FASTA sequences used for each allergen are available in Appendix 1.

Multiple sequence alignment, homology (A-RISC), and conserved motifs

The percentage identity and similarity calculations performed using SIAS for each protein family are presented in Appendix 2, while the visualization of conserved motifs generated with Jalview is included in Appendix 3.

Bet V 1 - Pathogenesis-related proteins PR-10

Classification: High risk of cross-reactivity.
PR-10 proteins exhibit sequence identity ranging from 67 to 72%, with similarity values close to 75%. These high values elevate the A-RISC index above the 0.70 threshold, suggesting a clinically significant risk of IgE-mediated cross-reactivity between *Ara h 8, Gly m 4*, and *Cic a 4*. These proteins are thermolabile and susceptible to digestive degradation, typically inducing mild symptoms such as oral allergy syndrome (OAS). Sensitization often originates from pollen exposure, and foods containing PR-10 proteins may trigger cross-reactive responses.

Several highly conserved regions were identified, such as “VEGNQGPGTIKKL” and “IDEALNGYMVSVVGGAALP” (residues ~40–90), which coincide with functional structural motifs, including a hydrophobic cavity involved in ligand transport and IgE recognition.

NsLTP: Nonspecific lipid transfer proteins

Classification: Low to moderate risk of cross-reactivity among legumes.

The A-RISC index analysis for nsLTPs reveals notable heterogeneity in sequence homology. While some protein pairs showed intermediate values (≥0.45), most comparisons fell below the 0.75 threshold, suggesting a low to moderate risk. Specific pairs, such as *Len c 3-Pis s 3* or *Cic a 3-Pis s 3*, demonstrated moderate homology and potential for partial cross-reactivity. *Ara h 9* showed low identity with other legume nsLTPs.

Figure 4. Conserved motif among NsLTP (Len c 3) (Jalview in Appendix 2).

Three conserved motifs were identified: “PC,” “CC,” and “CLK,” mostly between residues 40–90. These are part of the disulfide-stabilized core structure essential for IgE-binding regions. Additional motifs like “APC,” “YVRG,” “PGKCGV,” and “NPIY” define the immunogenic core of these stable proteins.

Glycinin (legumin, 11S globulin)

Classification: Low to moderate risk of cross-reactivity among legumes.

Table 2 Protein family classification.

AllFam family	Biochemical name	Function in plant	Heat stability	Digestive resistance	Representative allergens
Bet v 1 family	Pathogenesis-related protein, PR-10	Pathogenesis-related proteins involved in plant defense against pathogens.	Low	Low	Gly m 4 Ara h 8 Cic a 4
nsLTP family	Nonspecific lipid transfer protein	Lipid transfer between membranes; defence protein expressed in leaves, fruits, and seeds	High	High	Ara h 9 Ara h 16 Arah17 Lup an 3 Pis s 3 Len c 3 Cic a 3 Pha v 3
Prolamin superfamily	2S Albumin	Seed storage protein; provides sulfur-rich amino acids essential for seed germination.	High	High	Gly m 8 Lup an δ Ara h 2 Ara h 6 Ara h 7
Cupin superfamily	7S Vicilin	Seed storage; involved in nitrogen storage and antimicrobial defense.	High	High	Gly m 5 Ara h 1 Lup an 1 Pis s 1 Len c 1 Cic a 1
	11S Legumin	Major seed storage protein; abundant in legumes, supplies amino acids during germination.	High	High	Gly m 6 Ara h 3 Pis s 2 Cic a 6
Prolamin superfamily	2S Albumin	Seed storage protein; provides sulfur-rich amino acids essential for seed germination.	High	High	Gly m 8 Lup an δ Ara h 2 Ara h 6 Ara h 7
Profilin family	Profilin	Cytoskeletal regulation (actin binding); expressed in all eukaryotic cells	Low	Low	Gly m 3 Ara h 5
Oleosin family	Oleosin	Lipid body stabilization in oil-rich seeds	High	High	Ara h 10 Ara h 11 Ara h 14 Ara h 15
Defensin family	Defensin-like protein	Antimicrobial peptides; plant innate immunity	Variable	Variable	Gly m 2 Ara h 12 Ara h 13
Cyclophilin family	Cyclophilin	Protein folding (isomerase); involved in stress responses	Unknown	Unknown	Ara h 18

Adapted from Pomés et al. (2018)[8] and Sievers et al. (2011)[10]

Homology analysis showed an A-RISC value of 0.57 between *Ara h 3* and *Gly m 6*, suggesting a moderately high risk. Other comparisons showed lower similarity levels.

These proteins have two cupin domains (ß-barrels with surrounding α-helices) and form hexamers, conferring strong stability.

Cic a 4	1		
Gly m 4	0.72	1	
Ara h 8	0.69	0.71	1
	Cic a 4	Gly m 4	Ara h 8

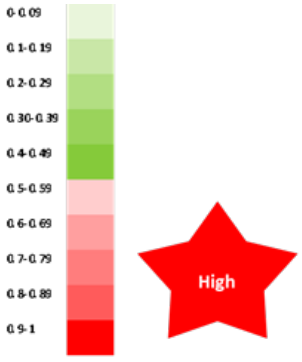


Figure 1 A-RISC among PR-10 protein sequences. A-RISC (Sequence homology index): A-RISC values were calculated to assess the degree of homology among allergens within the PR-10 family.

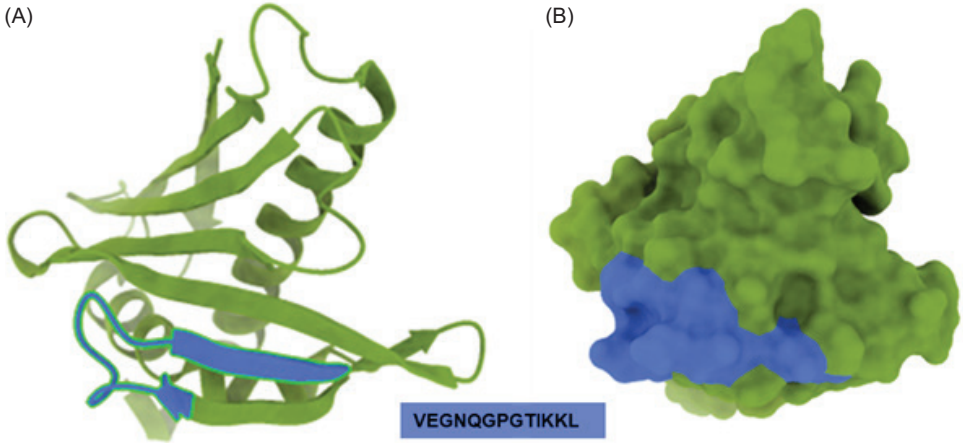


Figure 2 Conserved motif among PR-10 (Ara h 8). (A) Cartoon representation of the Ara h 8 protein, a member of the PR-10 family, highlighting in blue the conserved motif "VEGNQGPGTIKKL," located on one of the β-sheets that form part of the characteristic central hydrophobic cavity of the Bet v 1-like fold. (B) Surface representation of the same structure, where the *VEGNQGPGTIKKL* motif is clearly exposed, reinforcing its role as an immunologically relevant epitope. The structures were generated using UCSF ChimeraX (https://www.rbvi.ucsf.edu/chimerax/) based on the structural model of Ara h 8 (Jalview in Appendix 2).

Ara h 16	1							
Ara h 17	0.09	1						
Len c 3	0.08	0.5	1					
Pis s 3	0.09	0.51	0.64	1				
Cic a 3	0.13	0.5	0.65	0.63	1			
Lup an 3	0.07	0.45	0.53	0.5	0.5	1		
Pha v 3	0.04	0.48	0.5	0.46	0.5	0.6	1	
Ara h 9.0101	0.06	0.46	0.58	0.55	0.54	0.59	0.61	1
	Ara h 16	Ara h 17	Len c 3	Pis s 3	Cic a 3	Lup an 3	Pha v 3	Ara h 9.0101

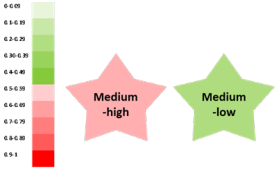


Figure 3 A-RISC among NsLTP protein sequences. A-RISC (Sequence homology index): A-RISC values were calculated to assess the degree of homology among allergens within the nsLTP family.

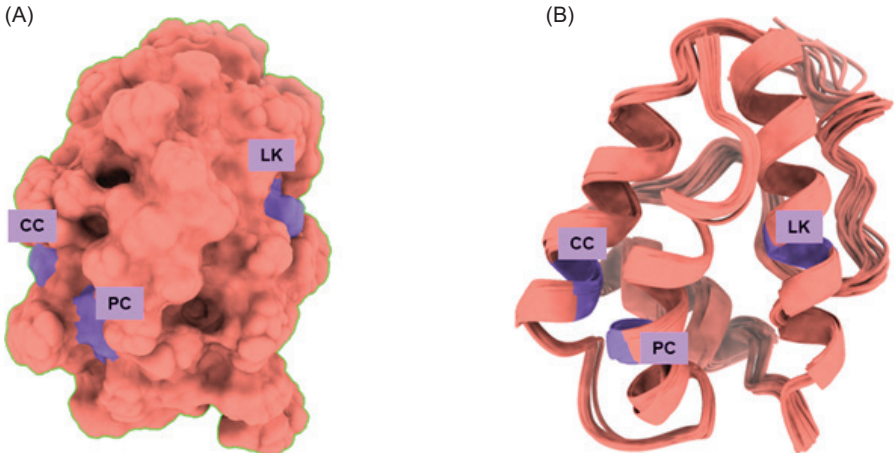


Figure 4 Conserved motif among NsLTP (Len c 3). (A) Surface representation of the Len c 3 protein (a lentil nsLTP), highlighting in violet the conserved motifs *PC*, *CC*, and *LK*, which are associated with structurally and immunologically relevant regions. (B) Cartoon representation of the same model, showing the spatial localization of these motifs within the compact fold characteristic of nsLTPs. These motifs correspond to surface-exposed or partially accessible residues that may contribute to structural stability, lipid transport, or immunological reactivity. The structures were generated using UCSF ChimeraX (https://www.rbvi.ucsf.edu/chimerax/) (Jalview in Appendix 2).

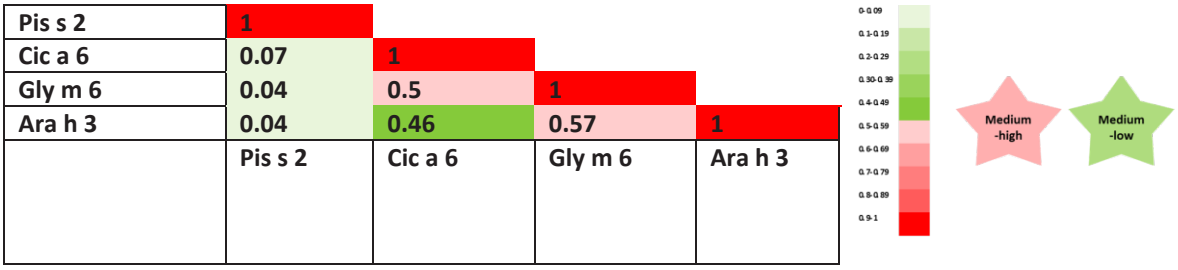


Figure 5 A-RISC among 11S globulin protein sequences. A-RISC (Sequence homology index): A-RISC values were calculated to assess the degree of homology among allergens within the 11S globulin.

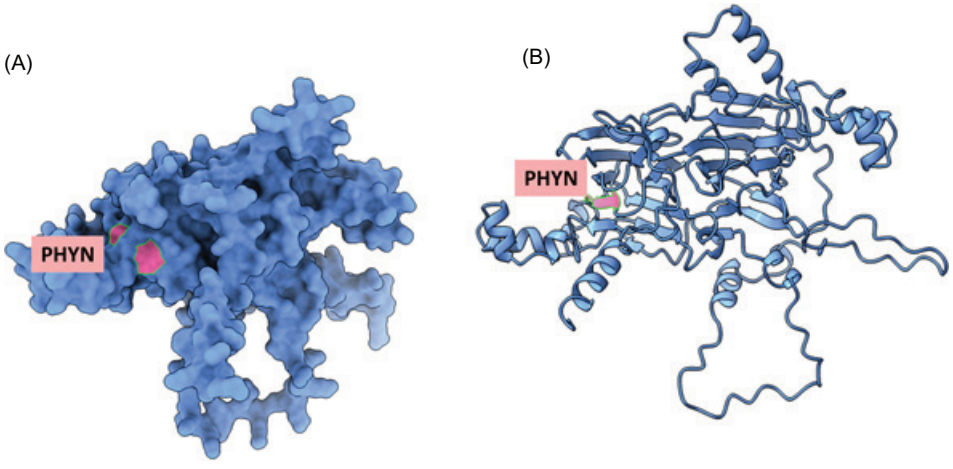


Figure 6 Conserved motif among 11S globulin (Gly m 6). (A) Surface representation of Gly m 6, an 11S globulin from soybean, highlighting in pink the conserved PHYN motif (Pro372-Asn375). This motif is localized on the surface, potentially accessible to IgE binding. (B) Cartoon representation of the same structure showing the PHYN motif situated within a ß-barrel characteristic of the cupin domain, which is a highly conserved structural feature among 11S globulins. The position and exposure of this motif suggest potential immunological relevance. Structures were generated using UCSF ChimeraX (https://www.rbvi.ucsf.edu/chimerax) (Jalview in Appendix 2).

Cic a 1	1					
Ara h 1	0.21	1				
Gly m 5	0.23	0.36	1			
Lup an 1	0.24	0.35	0.45	1		
Pis s 1	0.33	0.3	0.33	0.34	1	
Len c 1.0101	0.36	0.32	0.33	0.35	0.85	1
	Cic a 1	**Ara h 1**	**Gly m 5**	**Lup an 1**	**Pis s 1**	**Len c 1.0101**

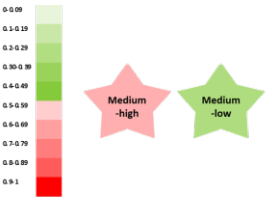


Figure 7 A-RISC among 7S Globulin protein sequences. A-RISC (Sequence homology index): A-RISC values were calculated to assess the degree of homology among allergens within the 7S Globulin.

In this study, the highly conserved "PHYN" motif was identified in the central regions of the sequences, forming part of the second cupin domain. This conservation, along with the moderate A-RISC values (~0.50–0.57), suggests a shared structural basis that may support the partial cross-reactivity observed clinically among these allergens.

Glycinin (Vicilins, 7S Globulin)

Classification: Low to moderate risk of cross-reactivity.

Vicilins are trimeric cupin superfamily proteins. While recognized as major allergens in peanut and soybean, sequence variability may limit cross-reactivity.

Conserved motifs include "VIV," "NLR," "ELVG," "GFGGINA," "FLAG," "DNVI," and "FPGS." These motifs are exposed on the trimer surface and are structurally positioned for IgE recognition, enhancing their immunologic relevance.

2S albumin

Classification: Low to moderately low cross-reactivity risk.

The A-RISC index values for 2S albumins among different legumes remain below the clinically relevant threshold of 0.50, indicating a low likelihood of IgE-mediated cross-reactivity between these species.

A conserved "ALQ" motif was located around position 140 in several allergens (*Ara h 2*, *Ara h 6*, *Ara h 7*, *Gly m 8*, *Lup an 6*). This motif may contribute to structural stability and surface exposure of IgE-binding sites. Although global homology is low, the preservation of these structural elements may explain clinical co-sensitization among legumes.

Summary of risk estimation

This table synthesizes the main findings of the A-RISC index across all allergenic families and species.

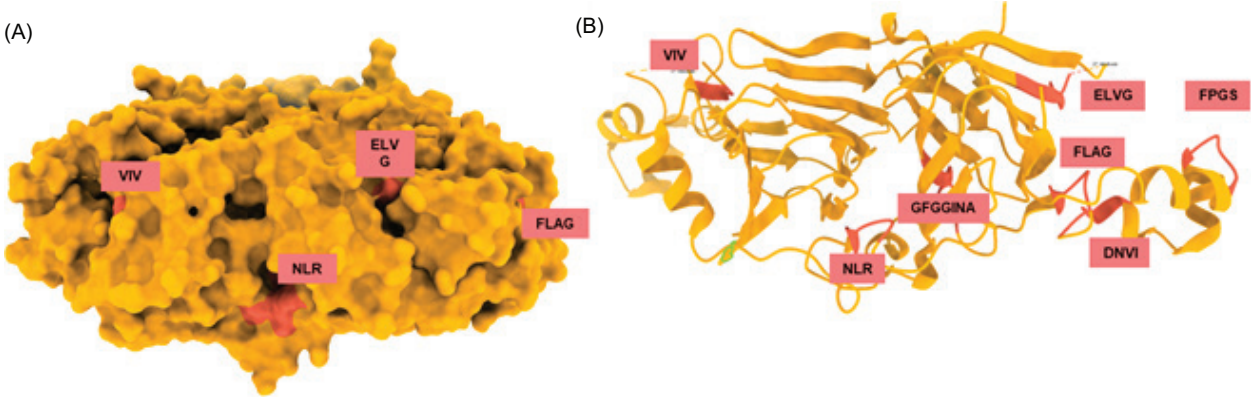


Figure 8 Conserved motif among 7S Globulin (Pis s 1). (A) Surface representation of Pis s 1 showing exposed conserved motifs on the protein surface, including VIV, ELVG, NLR, and FLAG (highlighted in pink). (B) Cartoon representation of the same structure showing both exposed and buried motifs. Motifs GFGGINA, DNVI, and FPGS are located internally and are not visible in the surface model, while VIV, NLR, ELVG, and FLAG are spatially accessible. These motifs may contribute to IgE-binding and structural stability. Structures were generated using UCSF ChimeraX (https://www.rbvi.ucsf.edu/chimerax) based on the crystallographic model of Pis s 1 (Jalview in Appendix 2).

Ara h 7	1				
Ara h 2	0.23	1			
Ara h 6	0.15	0.38	1		
Gly m 8	0.1	0.2	0.22	1	
Lup an δ	0.14	0.2	0.22	0.32	1
	Ara h 7	**Ara h 2**	**Ara h 6**	**Gly m 8**	**Lup an δ**

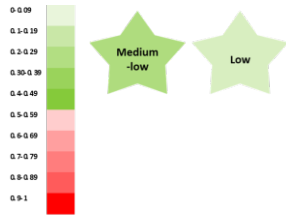


Figure 9 A-RISC among 2S Albumin protein sequences. A-RISC (Sequence homology index): A-RISC values were calculated to assess the degree of homology among allergens within the 2S Albumin.

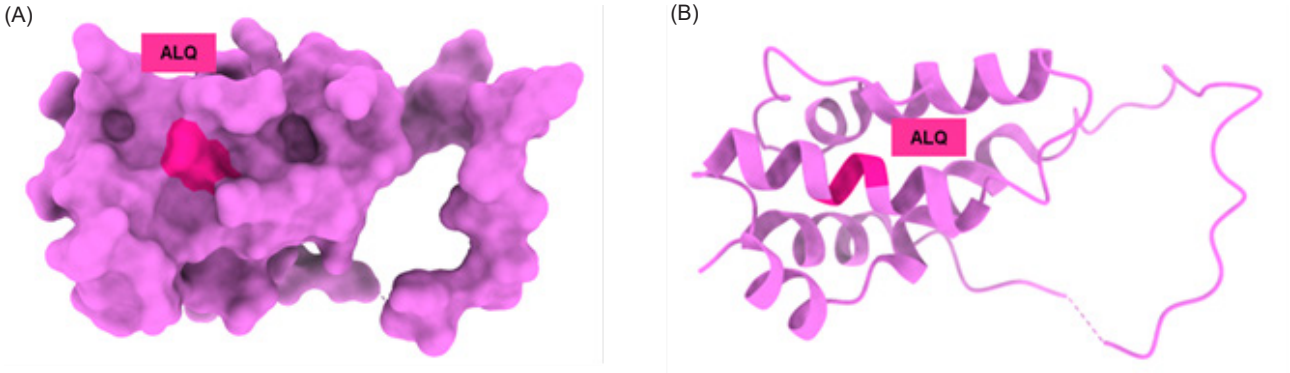


Figure 10 Conserved motif among 2S Albumin (Ara h 2). (A) Surface representation of Ara h 2 showing the conserved "ALQ" motif (highlighted in bright pink), located on a solvent-accessible region of the protein surface. (B) Cartoon representation of the same structure, illustrating the positioning of the "ALQ" motif within a short ß-strand embedded in the core α-helical bundle typical of the prolamin fold. This motif is conserved among several 2S albumins across legume species and may contribute to structural stability and allergenic potential. Visualizations were generated using UCSF ChimeraX (https://www.rbvi.ucsf.edu/chimerax) (Jalview in Appendix 2).

Table 3 Summary of in silico evaluation and estimated IgE cross-reactivity risk (A-RISC) across allergenic protein families from legumes.

Family	In silico analysis	A-RISC
PR-10 (Bet v 1-like)	High A-RISC (>0.70) between Ara h 8, Gly m 4, and Cic a 4. Sequence identity and similarity >67%. Conserved regions associated with IgE epitopes.	
nsLTP	Despite an overall low A-RISC (<0.50), it contains some conserved domains (**PC, CC, CLK**) potentially involved in structural stability or immunoreactivity.	
11S Globulinas	Conservation of the **"PHYN"** motif in central ß-barrel regions of the cupin domain; moderately conserved despite structural stability.	
7S Globulinas (Vicilinas)	Only one high pair (Len c 1-Pis s 1: A-RISC 0.85). Conservation of key peptide motifs (VIV, **FLAG, GFGGINA, etc.**) → "VIV," "NLR," "ELVG," "GFGGINA," "FLAG," "DNVI," "FPGS."	
2S Albúminas	Globally low A-RISC (<0.40), but presence of the conserved **"ALQ"** motif in 5 species.	

Mean values ± standard deviation are presented. Significant differences between groups were observed in baseline total IgE (pP = 0.031), baseline specific IgE to Pru p 3 (P = 0.046), and baseline specific IgG4 to Pru p 3 (P = 0.011).

Discussion

This study provides a comprehensive in silico comparative analysis of allergenic proteins from legumes, utilizing the A-RISC index and multiple sequence alignment to evaluate potential IgE cross-reactivity. Findings confirm that sequence homology alone does not fully predict cross-reactivity and must be complemented by structural motif analysis and epitope conservation.

Proteins such as Ara h 8 (peanut), Gly m 4 (soybean), and Cic a 4 (chickpea) exhibit high sequence similarity (67–72%), and conserved motifs such as "VEGNGGPGTIKKL," which overlap with immunogenic regions, are known to elicit IgE responses. PR-10 proteins are thermolabile and susceptible to digestive degradation, typically inducing mild symptoms such as oral allergy syndrome (OAS).[7] Sensitization commonly originates from pollen exposure, and subsequent ingestion of PR-10-containing foods can trigger cross-reactive responses, even in individuals who have not directly consumed all the legume sources involved.[11,12]

Ara h 9, considered a major nsLTP allergen in peanut,[13] exhibited low homology values with other legume nsLTPs such as lentil, pea, lupin, or bean, indicating a low probability of cross-reactivity between these species. This pattern aligns with previous literature, which has shown that although nsLTPs are highly stable and allergenic proteins, cross-reactivity between them depends on specific structural features and the individual sensitization profile of the patient. Multiple sequence alignments revealed conserved motifs such as "PC," "CC," and "CLK," predominantly located between residues ~40 and 90.[14] These motifs are essential for maintaining four disulfide bridges, which confer conformational stability and resistance to enzymatic digestion.[7,15] These findings align with structural studies by Salcedo et al. and Borges et al., showing that epitopes embedded within these stable structures are commonly recognized by IgE, even among nsLTPs with low overall sequence similarity.[14,15]

11S globulins are composed of two cupin-type domains, characterized by ß-barrels surrounded by α-helices. Their outer regions include conformational epitopes recognized by IgE. The cupin domain is functionally relevant and forms the structural basis shared across this protein family.[7] These proteins assemble into highly stable hexamers, which confer thermal and digestive resistance, properties that significantly enhance their allergenic potential.[5,16] In this study, the highly conserved "PHYN" motif was identified in the central regions of the sequences, forming part of the second cupin domain.[16,17] This conservation, along with the moderate A-RISC values (~0.50–0.57), suggests a shared structural basis that may support the partial cross-reactivity observed clinically among these allergens.

7S globulins (vicilins), such as Ara h 1 and Pis s 1, also belong to the cupin superfamily and typically assemble into trimers composed of 45–60 kDa subunits.[18] These proteins feature conserved ß-sheet domains forming solvent-exposed epitopes, including motifs like "VIV," "NLR," "ELVG," and "FPGS," which enhance IgE accessibility.[5] Their abundance and relative resistance to gastrointestinal digestion, along with the presence of both linear and conformational epitopes.[7,18] reinforce their role as clinically relevant allergens.

Finally, 2S albumins are low molecular weight proteins (12–15 kDa) belonging to the prolamin superfamily. These molecules are rich in cysteine residues and exhibit a compact tertiary structure stabilized by 4–5 disulfide bonds formed by a highly conserved cysteine pattern. These bonds are not typically exposed on the protein surface but instead provide structural support, maintaining the protein's tertiary conformation—especially the α-helical regions—and conferring exceptional resistance to thermal processing and enzymatic digestion.[19,20] As a result, 2S albumins persist during food processing and digestion, enhancing their allergenic potential in predisposed individuals.[21,22] Our multiple sequence alignment identified a conserved motif, "ALQ," located around position 140, shared by Ara h 2, Ara h 6, Ara h 7 (peanut), Gly m 8 (soybean), and Lup an 6 (lupin). This motif may play a role in stabilizing the conformational fold while also contributing to the surface exposure of linear IgE-binding epitopes. Although the global A-RISC values for 2S albumins were relatively low, suggesting limited cross-reactivity across legume species, the structural conservation of these motifs may underlie clinically observed co-sensitization, especially in individuals reactive to multiple legumes.[5,7]

While the A-RISC index provides a practical and scalable approach to estimate cross-reactivity likelihood among homologous allergens, it does not account for individual sensitization patterns, HLA variability, or patient-specific IgE repertoires. A clinical study by López-Torrejón et al.[16] found that individuals sensitized to lentil (Len c 1) often did not react to other vicilins, despite sequence and structural similarity, reinforcing that in silico predictions must be contextualized with clinical data.

This study is purely in silico in nature and does not introduce new algorithms or experimental validation. Therefore, its findings must not be interpreted as directly translatable to clinical diagnostics or management. Rather, this work serves as a theoretical basis for future in vitro or in vivo studies.

Moreover, the implications extend to food labeling and nutritional recommendations. International regulations, such as those from the EFSA and the US FDA, require the clear identification of allergens in processed foods.[6] Nevertheless, the implementation of food labeling continues to face significant challenges, particularly for plant-derived foods in developing countries, where allergen-specific data and diagnostic infrastructure may be limited.

Understanding which protein families exhibit higher cross-reactivity risk can inform both clinical practice and regulatory policies. However, further experimental studies and clinical validation are necessary to translate these in silico predictions into practical applications.

In conclusion, this in silico analysis of allergenic legume proteins revealed variable levels of cross-reactivity among structural families. PR-10 proteins showed high risk (A-RISC >0.70), supporting the use of component-resolved diagnostics in pollen-sensitized patients. Moderate risk was observed for 11S and 7S globulins, while nsLTPs and 2S albumins showed globally low A-RISC values but may still contribute to co-sensitization due to structural conservation.

These findings highlight the usefulness of the A-RISC index as a predictive tool and support molecular diagnostics to improve clinical interpretation and food labeling, particularly in cases of multiple sensitizations.

Acknowledgments

The author expresses sincere gratitude to Universidad El Bosque and Dra Abril, for its academic support and access to essential scientific and bioinformatics resources. Financial support from the Asociación Colombiana de Inmunología (ACOI) for publication is also gratefully acknowledged. Special thanks are extended to my family for their encouragement and unwavering support."

Ethical Approval

This study is an in silico analysis and did not involve any experiments on humans or animals.

Protection of Human and Animal Subjects

This study did not include any procedures involving human participants or animals.

Patient Data Protection

No patient data were used in this study. All analyses were performed using publicly available protein sequences.

Authors' Contributions

JSPV was responsible for the conceptualization, data collection, sequence alignment, bioinformatic analysis, figure and table generation, writing of the original draft, and final revision of the manuscript.

Conflicts of Interest

The author has no potential conflicts of interest to declare.

Funding

None.

References

1. Gupta RS, Warren CM, Smith BM, Blumenstock JA, Jiang J, Davis MM, et al. The public health impact of parent-reported childhood food allergies in the United States. Pediatrics. 2018;142(6):e20181235. https://doi.org/10.1542/peds.2018-1235
2. Sicherer SH, Sampson HA. Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. J Allergy Clin Immunol. 2014;133(2):291–307. https://doi.org/10.1016/j.jaci.2013.11.020
3. Sánchez A, Sánchez J, Cardona R. Results and limitations of epidemiological studies on food allergy. Focus on tropical countries. Rev Alerg Mex. 2019;66(1):9–17. https://doi.org/10.29262/ram.v66i1.340
4. Besler M, Steinhart H, Paschke A. Stability of food allergens and allergenicity of processed foods [Internet]. J Chromatography B. 2001. Available from: www.elsevier.com/locate/chromb
5. Breiteneder H, Mills ENC. Molecular properties of food allergens. J Allergy Clin Immunol. 2005;115(1):14–23. https://doi.org/10.1016/j.jaci.2004.10.022
6. Scientific Opinion on the evaluation of allergenic foods and food ingredients for labelling purposes. EFSA J. 2014;12(11):3894. https://doi.org/10.2903/j.efsa.2014.3894
7. Hoffmann-Sommergruber K, Hilger C, Santos A, De Las Vecillas L, Dramburg S. Molecular Allergology User's Guide 2.0. 2023;34 Suppl 28:e13854. https://doi.org/10.1111/pai.13854
8. Pomés A, Davies JM, Gadermaier G, Hilger C, Holzhauser T, Lidholm J, et al. WHO/IUIS Allergen Nomenclature: Providing a common language. Mol Immunol. 2018;100:3–13. https://doi.org/10.1016/j.molimm.2018.03.003
9. Chruszcz M, Kapingidza AB, Dolamore C, Kowal K. A robust method for the estimation and visualization of IgE cross-reactivity likelihood between allergens belonging to the same protein family. PLoS One. 2018;13(11):e0208276. https://doi.org/10.1371/journal.pone.0208276
10. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539. https://doi.org/10.1038/msb.2011.75
11. Ffitima Ferreira B, Hirtenlehner K, Jilek A, Godnik-Cvarfl Heimo Breiteneder J, Grimm R, Karin Hoffmann-Sommergruber I, et al. Dissection of immunoglobulin E and T lymphocyte reactivity of isoforms of the major birch pollen allergen Bet v 1: Potential use of hypoallergenic isoforms for immunotherapy. 1996:183(2):599–609. https://doi.org/10.1084/jem.183.2.599
12. Mittag D, Vieths S, Vogel L, Wagner-Loew D, Starke A, Hunziker P, et al. Birch pollen-related food allergy to legumes: Identification and characterization of the Bet v 1 homologue in mungbean (*Vigna radiata*), Vig r 1. Clin Exp Allergy. 2005;35(8):1049–55. https://doi.org/10.1111/j.1365-2222.2005.02309.x
13. Palladino C, Breiteneder H. Peanut allergens. Mol Immunol. 2018;100:58–70. https://doi.org/10.1016/j.molimm.2018.04.005
14. Salcedo G, Sánchez-Monge R, Barber D, Díaz-Perales A. Plant non-specific lipid transfer proteins: An interface between plant defence and human allergy. Biochim Biophys Acta. 2007;1771(6):781–91. https://doi.org/10.1016/j.bbalip.2007.01.001
15. Borges JP, Barre A, Culerrier R, Granier C, Didier A, Rougé P. Lipid transfer proteins from Rosaceae fruits share consensus epitopes responsible for their IgE-binding cross-reactivity. Biochem Biophys Res Commun. 2008;365(4):685–90. https://doi.org/10.1016/j.bbrc.2007.11.046
16. López-Torrejón G, Salcedo G, Martín-Esteban M, Díaz-Perales A, Pascual CY, Sánchez-Monge R. Len c 1, a major allergen and vicilin from lentil seeds: Protein isolation and cDNA cloning. J Allergy Clin Immunol. 2003;112(6):1208–15. https://doi.org/10.1016/j.jaci.2003.08.035
17. L'Hocine L, Boye JI. Allergenicity of soybean: New developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies. Crit Rev Food Sci Nutr. 2007;47(2):127–43. https://doi.org/10.1080/10408390600626487
18. Chruszcz M, Maleki SJ, Majorek KA, Demas M, Bublin M, Solberg R, et al. Structural and immunologic characterization of Ara h 1,

a major peanut allergen. J Biol Chem. 2011;286(45):39318–27. https://doi.org/10.1074/jbc.M111.270132
19. Moreno FJ, Clemente A. 2S Albumin storage proteins: What makes them food allergens? Open Biochem J. 2008;2:16–28. https://doi.org/10.2174/1874091X00802010016
20. Bueno-Díaz C, Martín-Pedraza L, Parrón J, Cuesta-Herranz J, Cabanillas B, Pastor Vargas C, Batanero E, Villalba M. Characterization of relevant biomarkers for the diagnosis of food allergies: An overview of the 2S albumin family. Foods. 2021;10(6):1235. https://doi.org/10.3390/foods10061235
21. Dreskin SC, Koppelman SJ, Andorf S, Nadeau KC, Kalra A, Braun W, et al. The importance of the 2S albumins for allergenicity and cross-reactivity of peanuts, tree nuts, and sesame seeds. J Allergy Clin Immunol. 2021;147(4):1154–63. https://doi.org/10.1016/j.jaci.2020.11.004
22. Sirvent S, Palomares O, Cuesta-Herranz J, Villalba M, Rodríguez R. Analysis of the structural and immunological stability of 2S albumin, nonspecific lipid transfer protein, and profilin allergens from mustard seeds. J Agric Food Chem. 2012;60(23):6011–8. https://doi.org/10.1021/jf300555h

Appendix 1

Prolamin superfamily

2S albumin (2S seed storage albumins family.)

>sp|P19594.2|2SS_SOYBN RecName: Full=2S seed storage albumin protein; AltName: Full=2S albumin; AltName: Full=GM2S-1; AltName: Full=Napin-type 2S albumin 3; Contains: RecName: Full=2S albumin small chain; AltName: Full=Aspartic acid-rich peptide; AltName: Full=Lunasin; Contains: RecName: Full=2S albumin large chain; AltName: Full=8 kDa methionine-rich protein; Short=8 kDa MRP; Flags: Precursor
MTKFTILLISLLFCIAHTCSASKWQHQQDSCRKQLQGVNLTPCEKHIMEKIQGRGDDDDDDDDDNHILRT
MRGRINYIRRNEGKDEDEEEEGHMQKCCTEMSELRSPKCQCKALQKIMENQSEELEEKQKKKMEKELINL
ATMCRFGPMIQCDLSSDD
>sp|Q99235.1|COND2_LUPAN RecName: Full=Conglutin delta 2; AltName: Allergen=Lup an delta-conglutin; Contains: RecName: Full=Conglutin delta-2 large chain; Contains: RecName: Full=Conglutin delta-2 small chain; Flags: Precursor
MAKLTILIALVAALVLVVHTSAFQSSKQSCKRQLQQVNLRHCENHIAQRIQQQQEEEEDHALKLRGIKHV
ILRHRSSQEYSEESEELDQCCEQLNELNSQRCQCRALQQIYESQSEQCEGSQQEQQLEQELEKLPRTCGF
GPLRRCDVNPDEE
>AAN77576.1 allergen Ara h 2.02 [Arachis hypogaea]
MAKLTILVALALFLLAAHASARQQWELQGDRRCQSQLERANLRPCEQHLMQKIQRDEDSYGRDPYSPSQD
PYSPSQDPDRRDPYSPSPYDRRGAGSSQHQERCCNELNEFENNQRCMCEALQQIMENQSDRLQGRQQEQQ
FKRELRNLPQQCGLRAPQRCDLEVESGGRDRY
>ABL14269.1 conglutin 8 [Arachis hypogaea]
MAKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLKPCEQHIMQRIMGEQEQYDSYDIRST
RSSDQQQRCCDELNEMENTQRCMGEALQQKMENQCDKLQDREMVEQFKRKLMDLAQQCNFRAPPRCDLDV
NGGRC
>AAD56719.1 allergen [Arachis hypogaea]
MMVKLSILVALLGALLVVASATRWDPDRGSRGSRWDAPSRGDDQCQRQLQRANLRPCEEHMRRRVEQEQE
QEQDEYPYSRRGSRGRQPGESDENQEQRCCNELNRFQNNQRCMCQALQQILQNQSFWVPAGQEPVASDGE
GAQELAPELRVQVTKPLRPL

Cupin superfamily

7S Vicilin

>sp|P11827.2|GLCAP_SOYBN RecName: Full=Beta-conglycinin alpha' subunit; Short=CG-alpha'-1; AltName: Full=Beta-conglycinin alpha prime subunit; AltName: Allergen=Gly m 5; Flags: Precursor
MMRARFPLLLLGVVFLASVSVSFGIAYWEKQNPSHNKCLRSCNSEKDSYRNQACHARCNLLKVEEEEECE
EGQIPRPRPQHPERERQQHGEKEEDEGEQPRPFPFPRPRQPHQEEEHEQKEEHEWHRKEEKHGGKGSEEE
QDEREHPRPHQPHQKEEEKHEWQHKQEKHQGKESEEEEEDQDEDEEQDKESQESEGSESQREPRRHKNKN
PFHFNSKRFQTLFKNQYGHVRVLQRFNKRSQQLQNLRDYRILEFNSKPNTLLLPHHADADYLIVILNGTA
ILTLVNNDDRDSYNLQSGDALRVPAGTTYYVVNPDNDENLRMITLAIPVNKPGRFESFFLSSTQAQQSYL
QGFSKNILEASYDTKFEEINKVLFGREEGQQQGEERLQESVIVEISKKQIRELSKHAKSSSRKTISSEDK
PFNLRSRDPIYSNKLGKLFEITPEKNPQLRDLDVFLSVVDMNEGALFLPHFNSKAIVVLVINEGEANIEL
VGIKEQQQRQQQEEQPLEVRKYRAELSEQDIFVIPAGYPVVVNATSDLNFFAFGINAENNQRNFLAGSKD
NVISQIPSQVQELAFPGSAKDIENLIKSQSESYFVDAQPQQKEEGNKGRKGPLSSILRAFY
>sp|Q43626|Q43626_PEA Vicilin 47kD protein
MAATPIKPLMLLAIAFLASVCVSSRSDQENPFIFKSNRFQTLYENENGHIRLLQKFDKRSKIFENLQNYR
LLEYKSKPHTLFLPQYTDADFILVVLSGKATLTVLKSNDRNSFNLERGDAIKLPAGTIAYLANRDDNEDL
RVLDLAIPVNKPGQLQSFLLSGTQNQPSLLSGFSKNILEAAFNTNYEEIEKVLLEQQEQEPQHRRSLKDR
RQEINEENVIVKVSREQIEELSKNAKSSSKKSVSSESGPFNLRSRNPIYSNKFGKFFEITPEKNQQLQDL
DIFVNSVDIKEGSLLLPNYNSRAIVIVTVTEGKGDFELVGQRNENQGKENDKEEEQEEETSKQVQLYRAK
LSPGDVFVIPAGHPVAINASSDLNLIGFGINAENNERNFLAGEEDNVISQVERPVKELAFPGSSHEVDRL
LKNQKQSYFANAQPLQRE
>sp|F5B8V9.1|CONB1_LUPAN RecName: Full=Conglutin beta 1; AltName: Allergen=Lup an 1; Flags: Precursor
MAKMRVRLPMLILLLGVVFLLAASIGIAYGEKDFTKNPPKEREEEEHEPRQQPRPRQQEEQEREHRREEK
HDGEPSRGRSQSEESQEEEHERRREHHRREREQEQQPRPQRRQEEEEEEEEWQPRRQRPQSRREEREEREQ
EQGSSSGSQRGSGDERRQHRERRVHREEREQEQDSRSDSRRQRNPYHFSSNRFQTYYRNRNGQIRVLERF
NQRTNRLENLQNYRIIEFQSKPNTLILPKHSDADFILVVLNGRATITIVNPDKRQVYNLEQGDALRLPAG
TTSYILNPDDNQNLRVAKLAIPINNPGKLYDFYPSTTKDQQSYFSGFSKNTLEATFNTRYEEIERVLLGD
DELQENEKQRRGQEQSHQDEGVIVRVSKKQIQELRKHAQSSSGEGKPSESGPFNLRSNKPIYSNKFGNFY
EITPDINPQFQDLNISLTFTEINEGALLLPHYNSKAIFIVVVDEGEGNYELVGIRDQQRQQDEQEEEYEQ

GEEEVRRYSDKLSKGDVFIIPAGHPLSINASSNLRLLGFGINANENQRNFLAGSEDNVIKQLDREVKELT
FPGSIEDVERLIKNQQQSYFANAQPQQQQQREKEGRRGRRGPISSILNALY
>sp|Q84UI1|Q84UI1_LENCU Allergen Len c 1.0101
SRSDQENPFIFKSNRFQTIYENENGHIRLLQRFDKRSKIFENLQNYRLLEYKSKPHTIFLPQFTDADFIL
VVLSGKAILTVLNSNDRNSFNLERGDTIKLPAGTIAYLANRDDNEDLRVLDLAIPVNRPGQLQSFLLSGT
QNQPSFLSGFSKNILEAAFNTEYEEIEKVLLEEQEQKSQHRRSLRDKRQEITNEDVIVKVSREQIEELSK
NAKSSSKKSVSSESEPFNLRSRNPIYSNKFGKFFEITPEKNPQLQDLDIFVNSVEIKEGSLLLPNYNSRA
IVIVTVNEGKGDFELVGQRNENQQEQREENDEEEGQEEETTKQVQRYRARLSPGDVLVIPAGHPVAINAS
SDLNLIGFGINAKNNQRNFLAGEEDNVISQIQRPVKELAFPGSSREVDRLLTNQKQSHFANAQPLQIE
>CAA36188.1 provicilin precursor [Cicer arietinum]
MIVRFSLPDNENDLKLTRSINRDGEILIPKIFIIISVSQISNGASREFDGISSLKVEVFLSLGFNTVSIA
LHLGLQDGSRHHCVVEERGCEVLSYFLQTVVLEVLKLLRTFVEPLEKTNVTVFVLEKSLKVVRLKEKRIL
LIGISHKLRPRKQFLSSTKSGNRALIAILMIEFLLSFRIDDEIERVLLEEQEQKPKQRRGHKDRQQSQSQ
SQQEADVIVKISREQIEELSKNAKSSSKKSVSSESEPFNLRSRNPIYSNKYGNFFEITPEKNPQLQDLDI
SLNSVEINEGSLLLPHFNSRATVILVVNEGKGEVELVGLRNENEQENKKEDEEEEEDRKVQVQRFQSRLS
SGDVVVIPATHPFSINASSDLFLLGFGINAQNNQRNFLAGEEDNVISQIQRPVKEVAFPGSAEEVDRLLK
NQRQSHFANAQPQQKDEESQKIRIPLSSILGGF
>ACF22884.1 main allergen Ara h1 [Arachis hypogaea]
MRGRVSPLMLLLGILVLASVSATQAKSPYRKTENPCAQRCLQSCQQEPDDLKQKACESRCTKLEYDPRCV
YDTGATNQRHPPGERTRGRQPGDYDDDRRQPRREEGGRWGPAEPREREREEDWRQPREDWRRPSHQQPRK
IRPEGREGEQEWGTPGSEVREETSRNNPFYFPSRRFSTRYGNQNGRIRVLQRFDQRSKQFQNLQNHRIVQ
IEARPNTLVLPKHADADNILVIQQGQATVTVANGNNRKSFNLDEGHALRIPSGFISYILNRHDNQNLRVA
KISMPVNTPGQFEDFFPASSRDQSSYLQGFSRNTLEAAFNAEFNEIRRVLLEENAGGEQEERGQRRRSTR
SSDNEGVIVKVSKEHVQELTKHAKSVSKKGSEEEDITNPINLRDGEPDLSNNFGRLFEVKPDKKNPQLQD
LDMMLTCVEIKEGALMLPHFNSKAMVIVVVNKGTGNLELVAVRKEQQQRGRREQEWEEEEDEEEEGSNR
EVRRYTARLKEGDVFIMPAAHPVAINASSELHLLGFGINAENNHRIFLAGDKDNVIDQIEKQAKDLAFPG
SGEQVEKLIKNQRESHFVSARPQSQSPSSPEKEDQEEENQGGKGPLLSILKAFN

Glycinin (legumin, 11S globulin)
>sp|P04776.2|GLYG1_SOYBN RecName: Full=Glycinin G1; Short=Glycinin 11S G1; Short=Glycinin A1aB1b; AltName: Allergen=Gly m 6; Contains: RecName: Full=Glycinin A1a subunit; Short=Glycinin acidic 1a subunit; Contains: RecName: Full=Glycinin Bx subunit; Short=Glycinin basic x subunit; AltName: Full=Glycinin B1b subunit; Short=Glycinin basic 1b subunit; Flags: Precursor
MAKLVFSLCFLLFSGCCFAFSSREQPQQNECQIQKLNALKPDNRIESEGGLIETWNPNNKPFQCAGVALS
RCTLNRNALRRPSYTNGPQEIYIQQGKGIFGMIYPGCPSTFEEPQQPQQRGQSSRPQDRHQKIYNFREGD
LIAVPTGVAWWMYNNEDTPVVAVSIIDTNSLENQLDQMPRRFYLAGNQEQEFLKYQQEQGGHQSQKGKHQ
QEEENEGGSILSGFTLEFLEHAFSVDKQIAKNLQGENEGEDKGAIVTVKGGLSVIKPPTDEQQQRPQEEE
EEEEDEKPQCKGKDKHCQRPRGSQSKSRRNGIDETICTMRLRHNIGQTSSPDIYNPQAGSVTTATSLDFP
ALSWLRLSAEFGSLRKNAMFVPHYNLNANSIIYALNGRALIQVVNCNGERVFDGELQEGRVLIVPQNFVV
AARSQSDNFEYVSFKTNDTPMIGTLAGANSLLNALPEEVIQHTFNLKSQQARQIKNNNPFKFLVPPQESQ

KRAVA
>sp|P13915.1|CVCA_PEA RecName: Full=Convicilin; Flags: Precursor

MATTVKSRFPLLLFLGIIFLASVCVTYANYDEGSETRVPGQRERGRQEGEKEEKRHGEWRPSYEKEEHEE
EKQKYRYQREKKEQKEVQPGRERWEREEDEEQVEEEWRGSQRREDPEERARLRHREERTKRDRRHQREGE
EEERSSESQEHRNPFLFKSNKFLTLFENENGHIRRLQRFDKRSDLFENLQNYRLVEYRAKPHTIFLPQHI
DADLILVVLNGKAILTVLSPNDRNSYNLERGDTIKIPAGTTSYLVNQDDEEDLRVVDFVIPVNRPGKFEA
FGLSENKNQYLRGFSKNILEASLNTKYETIEKVLLEEQEKKPQQLRDRKRTQQGEERDAIIKVSREQIEE
LRKLAKSSSKKSLPSEFEPFNLRSHKPEYSNKFGKLFEITPEKKYPQLQDLDILVSCVEINKGALMLPHY
NSRAIVVLLVNEGKGNLELLGLKNEQQEREDRKERNNEVQRYEARLSPGDVVIIPAGHPVAISASSNLNL
LGFGINAKNNQRNFLSGSDDNVISQIENPVKELTFPGSSQEVNRLIKNQKQSHFASAEPEQKEEESQRKR
SPLSSVLDSFY
>CAB60140.1 legumin, alpha and beta subunit [Cicer arietinum]
MAKLLALSLSFCFLLFGTCFALRDQPQQNECQLEHLNALKPDNRIKSEGGLIETWNPSNKQFACAGVALS
RATLQPNSLLQTFLHQRSPEIFIQQGNGYFGMVFPGCVETFEEPRESEQGEGSKFSDSHQKVNRFREGDI
IAVPTGVVFWMFNDQDTPVIAVSLIDTSSFQNQLDQMPRRFYLAGNHEQEFLRYQQEGSEEEENEGGNIF
SGFKRDFLEDALNVNRRIVNKLQGRNEDEEKGAIVKVKGGLSITTPPEKEPRQKRGSRQEEDEDEDEKRQ
PHRHSRQDEDEDEKRQPHHHSRGGSKSQRDNGFEETICTARLHQNIGSSSSPDIYNPQAGRIKTVTSFDL
QALRFLKLSAEFGSLHKNAMFVPHYNLNANSILYALKGRARLLYALNCKGNSVFDGELEAGRALIVPQNF
AIAAKSLSDRFSYVAFKTNDRALINVCQKKLLQLLSIWKEMRPGSSSSTAPFHFLFHPAVTQTTKQQLDL

VPNQYE

>tr|Q647H3|Q647H3_ARAHY Arachin Ahy-2 OS=Arachis hypogaea OX=3818 PE=2 SV=1
MAKLLALSVCFCFLVLGASSISFRQQPEENACQFQRLNAQRPDNRLESEGGYIETWNPNN
QEFECAGVALSRLVLRRNALRRPFYSNAPQEIFIQQGRGYFGLIFPGCPSTYEEPAQQGR
RHQSQRAPRRFEGEDQSQQQQQDSHQKVRRFDEGDLIAVPTGVALWMYNDHDTDVVAVSL
TDTNNNDNQLDQFPRRFNLAGNHEQEFLRYQQQSRRRSLPYSPYSPQSQPRQEEREFSPR
GQHSRRERAGQEQENEGGNIFSGFTPEFLAQAFQVDDRQILQNLRGENESDEQGAIVTVR
GGLRILSPDRKRRQQYERPDEEEEYDEDEYEYDEEERQQDRRRGRGSRGRGNGIEETICT
ASVKKNIGRNRSPDIYNPQAGSLKTANDLNLLILRWLGLSAEYGNLYRNALFVPHYNTNA
HSIIYALRGRAHVQVVDSNGNRVYDEELQEGHVLVVPQNFAVAGKSQSDNFEYVAFKTDS
RPSIANLAGENSIIDNLPEEVVANSYGLPREQARQLKNNNPFKFFVPPSQQSLGAVA

NsLTP

>sp|A0A158V755.1|NLTP2_PEA RecName: Full=Non-specific lipid-transfer protein 2; Short=PsLTP2; Flags: Precursor
MATSMKLACVALVMCMVVIAPMAEAALSCGTVSGDLAPCLTYLQAPNNASPPPPCCAGVKKLLGAATTTP
DRQAACNCLKSAAGSISRLNTNNAAALPGKCGVSIPYKISTSTNCNTIKF
>OIW00903.1 hypothetical protein TanjilG_19844 [Lupinus angustifolius]
MASIKVACVVLMCMAVVAAPIAQAITCGQVVGNLAPCITYLRSGGAVPPSCCGGVKSLVSSAQTTADKRT
VCGCLKSAVGAIPNYNDANAAALPGKCGVSVPYKISVSTNCATYVLFSLF
>AAX35806.1 lipid transfer protein 1 precursor [Lens culinaris]
MASLRVSCLVALMCMVVISAPMAEAAISCGTVSGALVPCLTYLKGGPGPSPQCCGGVKRLNGAARTTIDR
RAACNCLKSSAGSISGLKPGNVATLPGKCGVRLPYTISTSTNCNTIRF
>ADC80502.1 non-specific lipid transfer protein 1a precursor [Phaseolus vulgaris]
MASVKFACVVVLCMVVVGAHTAQGMTCGQVQSNLVPCVTFLQNGGFVPAGCCNGVRNIMNSARSTADRRG
ICNCLKTAAGAVRGLNPNNAQALPGKCGVNIPYKISTSTNCASIN
>CAA05771.1 lipid transfer protein [Cicer arietinum]
MASMKVVCVALIMCIVIAPMAESAITCGRVDTALAPCLGYLQGGPGPSAQCCGGVRNLNSAAVTTPDRQA
ACNCLKSAAGSISRLNANNAAALPGKCVVNIPYKISTSTNCATIRV
>ABX56711.1 LTP isoallergen 1 precursor [Arachis hypogaea]
MASLKFAFVMLVCMAMVGAPMVNAISCGQVNSALAPCIPFLTKGGAPPPACCSGVRGLLGALRTTADRQA
ACNCLKAAAGSLRGLNQGNAAALPGRCGVSIPYKISTSTNCATIKF
>RYR60032.1 hypothetical protein Ahy_A04g017131 [Arachis hypogaea]
MMMKKVCAVLVVALMVLVEVAPMAEAVTCTPTELSPCLGAITGGSPPSSVCCQKLRAQKPCLCNYIKNPA
LRTYVNSPGARRVASSCGVPLPSC
>RYR27153.1 hypothetical protein Ahy_B02g061489 [Arachis hypogaea]
MAKLAPCVVLMLCMAIVGAPIAKAAIQCSFVTKSIAPCFGYLKSGGTVSGPCCSGIQNINGTAKTTSDRQ
AVCNCLKSAAASLGSQINPNNAASLPGKCGVSIPYKISTSINCSSIK

Pathogenesis-related protein, PR-10

>sp|P26987.1|SAM22_SOYBN RecName: Full=Stress-induced protein SAM22; AltName: Full=Pathogenesis-related protein 10; AltName: Full=Starvation-associated message 22; AltName: Allergen=Gly m 4
MGVFTFEDEINSPVAPATLYKALVTDADNVIPKALDSFKSVENVEGNGGPGTIKKITFLEDGETKFVLHK
IESIDEANLGYSYSVVGGAALPDTAEKITFDSKLVAGPNGGSAGKLTVKYETKGDAEPNQDELKTGKAKA
DALFKAIEAYLLAHPDYN
>CAA56142.1 pathogenesis related protein [Cicer arietinum]
MGVFTFEQETASTVPPAKLYKAMVKDADVIIPKAVDAIKTVETVEGNGGPGTIKKLTFVEGGQTLYVLHK
IEAIDEANLGYNYSIVGGAGLSETVERYHFEAKLCEGPNGGSIGKVSVKYQTKGDAKPNEKEVQEGKAKG
DALFKAIEGYVLANPNYN
>ACD39391.1 pathogenesis-related class 10 protein [Arachis hypogaea]
MGVFTFEDEITSTLPPAKLYNALKDADTITPKIIDDVKSVEIVEGNGGPGTIKKLTIVEDGETKFILHKV
ESIDEANYAYNYSVVGGVALPPTAEKITFETKLVEGPNGGSIGKLTLKYHTKGDAKPDEEELKKGKAKGEGLFRAIEGYALANPSQY

Appendix 2
Calculation of Identity and Similarity Using Sias

IDENTITY: PR-10

Cic a 4	100		
Gly m 4	68.35	100	
Ara h 8	67.51	70.7	100
	Cic a 4	**Gly m 4**	**Ara h 8**

Sequence identity (%) among PR-10 protein sequences.

SIMILARITY : PR-10

Cic a 4	100		
Gly m 4	77.2	100	
Ara h 8	77.07	78.98	100
	Cic a 4	**Gly m 4**	**Ara h 8**

Sequence similarity (%) among PR-10 protein sequences.

IDENTITY: nsLTP

Ara h 16	100							
Ara h 17	27.65	100						
Len c 3	30.85	49.57	100					
Pis s 3	31.91	54.7	67.79	100				
Cic a 3	38.29	54.31	68.96	71.55	100			
Lup an 3	28.72	49.57	55.08	55.83	57.75	100		
Pha v 3	29.78	50.43	51.3	54.78	56.52	59.13	100	
Ara h 9.0101	27.65	52.58	59.48	63.79	60.34	58.62	60	100
	Ara h 16	**Ara h 17**	**Len c 3**	**Pis s 3**	**Cic a 3**	**Lup an 3**	**Pha v 3**	**Ara h 9.0101**

Sequence identity (%) among NsLTP protein sequences.

SIMILARITY: nsLTP

Ara h 16	100							
Ara h 17	41.48	100						
Len c 3	45.74	61.53	100					
Pis s 3	47.87	64.1	74.57	100				
Cic a 3	51.06	62.06	74.13	76.72	100			
Lup an 3	40.42	58.97	62.71	63.33	63.79	100		
Pha v 3	43.61	60	64.34	63.47	65.21	70.43	100	
Ara h 9.0101	41.48	60.34	68.1	70.68	68.1	67.24	70.43	100
	Ara h 16	**Ara h 17**	**Len c 3**	**Pis s 3**	**Cic a 3**	**Lup an 3**	**Pha v 3**	**Ara h 9.0101**

Sequence similarity (%) among NsLTP protein sequences.

IDENTITY: 11S Globulin

Pis s 2	100			
Cic a 6	15.72	100		
Gly m 6	17.37	55.35	100	
Ara h 3	16.75	51.2	60.2	100
	Pis s 2	**Cic a 6**	**Gly m 6**	**Ara h 3**

Sequence identity (%) among 11S Globulin protein sequences.

SIMILARITY: 11S Legumin

Pis s 2	100			
Cic a 6	23.38	100		
Gly m 6	26.46	63.23	100	
Ara h 3	26.44	58.87	69.09	100
	Pis s 2	**Cic a 6**	**Gly m 6**	**Ara h 3**

Sequence similarity (%) among 11S Globulin protein sequences.

IDENTITY: 7S Globulin

Cic a 1	100					
Ara h 1	36.86	100				
Gly m 5	37.74	42.34	100			
Lup an 1	40.83	42.88	49.09	100		
Pis s 1	47.26	48.85	52.28	54.56	100	
Len c 1.0101	51.67	52.87	54.3	56.69	90.19	100
	Cic a 1	**Ara h 1**	**Gly m 5**	**Lup an 1**	**Pis s 1**	**Len c 1.0101**

Sequence identity (%) among 7S vicilin protein sequences.

SIMILARITY: 7S Globulin

Len c 1.0101	60.76	64.11	66.02	66.02	93.06	100
	Cic a 1	**Ara h 1**	**Gly m 5**	**Lup an 1**	**Pis s 1**	**Len c 1.0101**

Sequence similarity (%) among 7S Vicilin protein sequences.

IDENTITY: 2S Albumin

Ara h 7	100				
Ara h 2	38.12	100			
Ara h 6	34.48	54.48	100		
Gly m 8	25.94	35.44	32.41	100	
Lup an **δ**	34.64	35.94	34.48	39.21	100
	Ara h 7	**Ara h 2**	**Ara h 6**	**Gly m 8**	**Lup an δ**

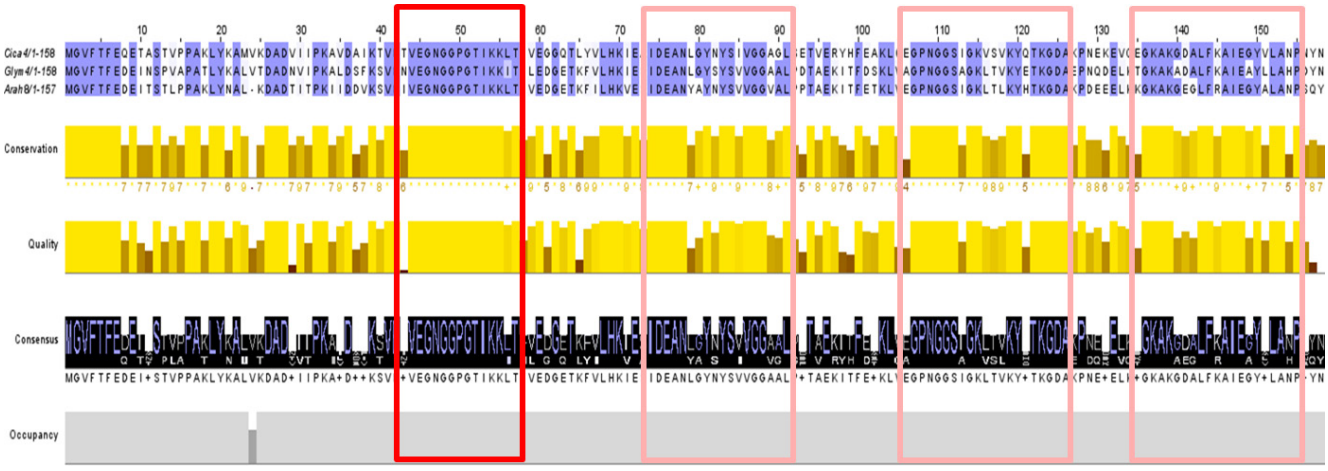
Sequence identity (%) among 2S Albumin protein sequences.

SIMILARITY: 2S Albumin

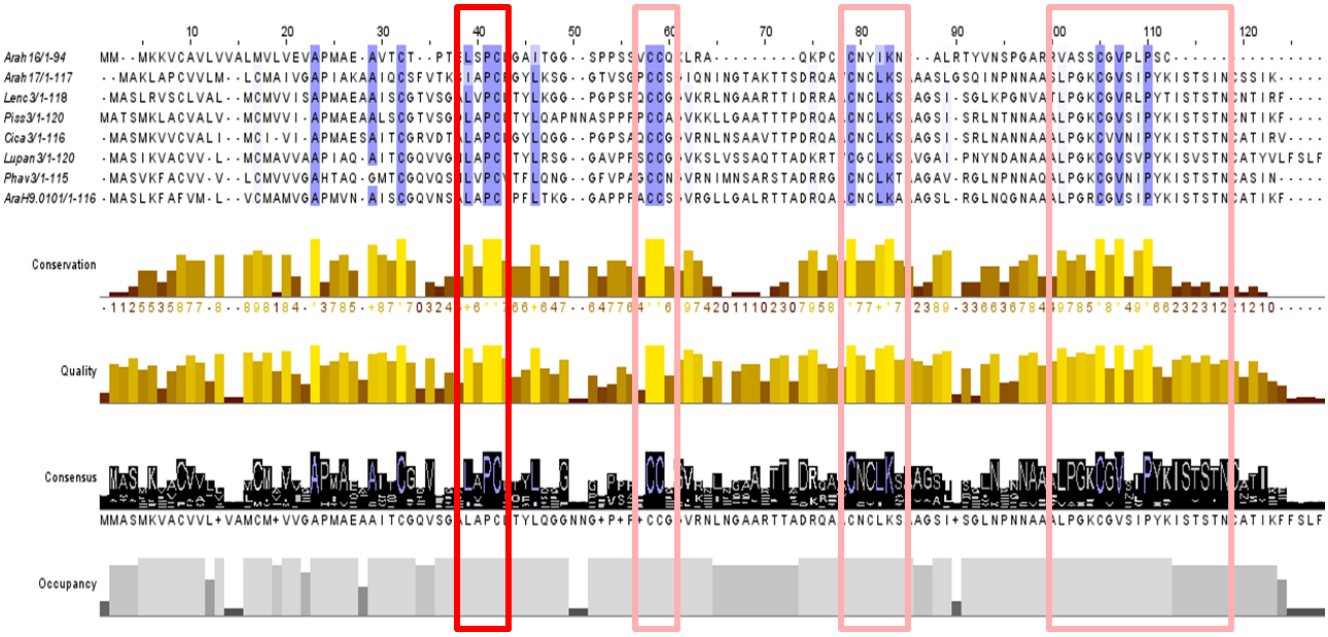
Ara h 7	100				
Ara h 2	44.37	100			
Ara h 6	39.31	61.37	100		
Gly m 8	36.7	41.13	42.06	100	
Lup an **δ**	38.56	42.48	42.06	48.36	100
	Ara h 7	**Ara h 2**	**Ara h 6**	**Gly m 8**	**Lup an δ**

Sequence similarity (%) among 2S Albumin protein sequences.

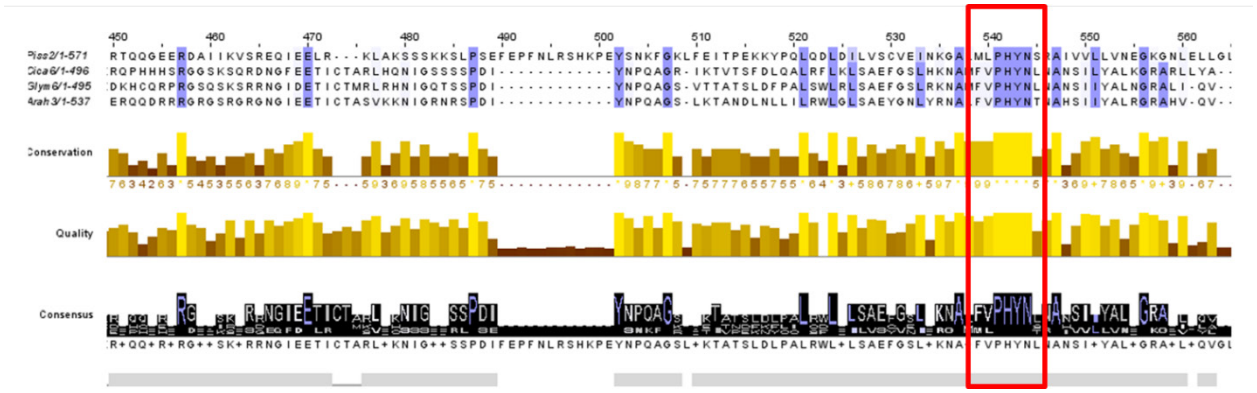
Appendix 3
Multiple sequence alignment and conserved motifs visualization using Jalview



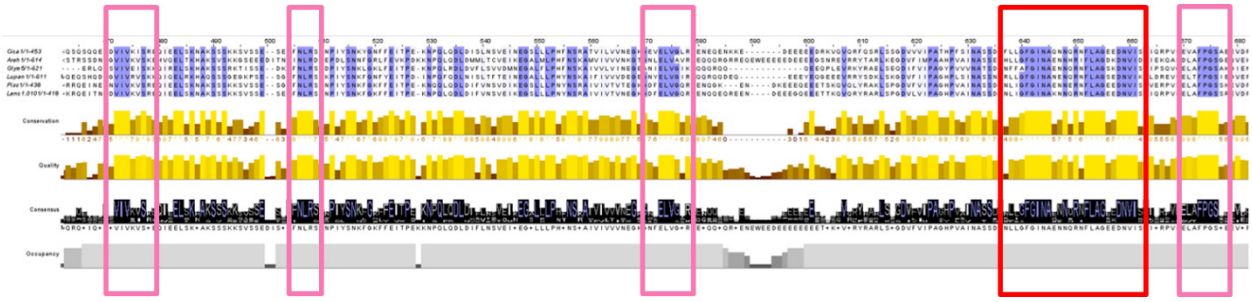
Multiple sequence alignment among PR-10.



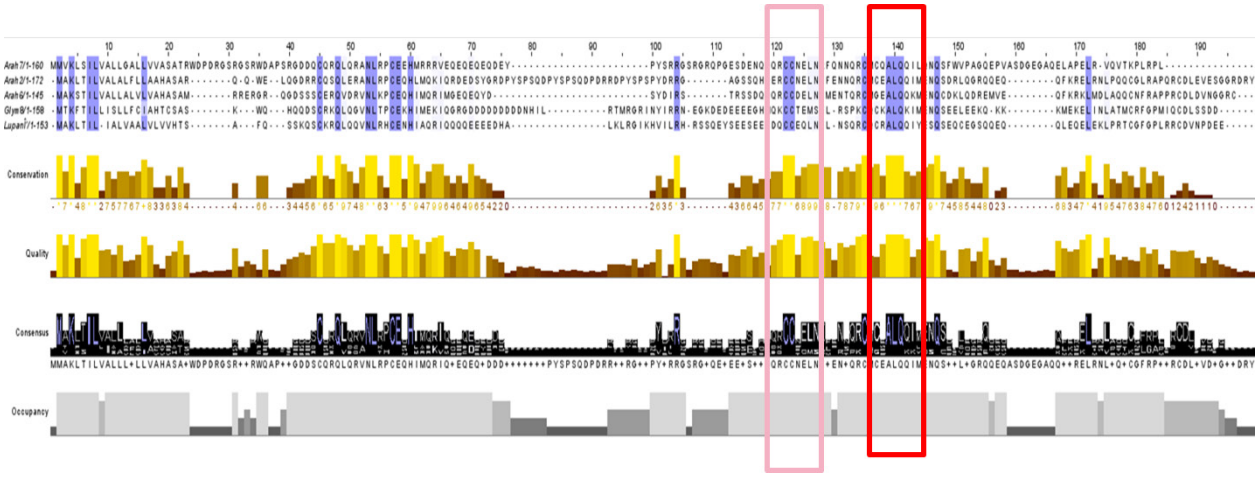
Multiple sequence alignment among nsLTP.



Multiple sequence alignment among 11S Legumin.



Multiple sequence alignment among 7S Vicilin.



Multiple sequence alignment among 2S Albumin.