Chapter

Comparative Analysis of Molecular Allergy Features of Seed Proteins from Soybean (Glycine max) and Other Legumes Extensively Used for Food

Andrea Roman-Mateo, Esther Rodriguez-de Haro, Jose M. Berral-Hens, Sonia Morales-Santana and Jose C. Jimenez-Lopez

Abstract

Food allergies due to eating habits, pollution, and other factors are a growing problem in Western nations as well as developing countries. Symptoms of food allergies include changes in the respiratory and digestive systems. Legumes are a potential solution to the enormous demands for healthy, nutritive, and sustainable food. However, legumes also contain families of proteins that can cause food allergies. Some of these legumes include peanut, pea, chickpea, soy, and lupine. It has been shown that processing can alter the allergenicity of legumes since thermic and enzymatic resistance can affect these properties. Cross-reactivity (CR) is an allergy feature of some allergen proteins when the immune system recognizes part of the common share sequences (epitopes) in these allergic proteins. The research about molecular allergy includes comparisons of immunoglobulin E (IgE) and T-cell epitopes, assessment of three-dimensional structure and comparison of secondary structure elements, post-transduction modifications analysis by bioinformatic approach, and post-transduction modifications affecting epitopes properties may facilitate molecular tools to predict protein allergic behavior establishing prevention measurements that could promote the use of legumes and other seeds. This chapter provides an overview of the structural features of the main allergen proteins from legumes and their allergenic potential.

Keywords: food allergy, cross allergenicity, legumes, allergen proteins, soy, lupine

1. Introduction

Legumes are dicotyledon plants in the order Fabales and the family Fabacea. They produce fruit contained in pods and filled with seeds. In this chapter, we discuss three

species of legumes in the genus *Lupinus* (*Lupinus albus*, *L. angustiflora*, and *Lupinus luteus*) and the most common allergenic species of the family Papilionaceae, including soja *Glycine max*; *Arachis hypogaea*, *A. duraensis*, and *A. ipaensis*; lentil (*Lens culinaris*); pea (*Pisum sativum*); and chickpea (*Cicer arietinum*).

A food allergy is an immune system reaction that occurs after eating certain types of food. Symptoms are variable and can be caused even by small amounts of allergenic proteins, leading to hives, swollen airways, and digestive problems. Food allergies are a growing concern worldwide. This increase is suspected to be related with industrial production, pollution, additives, and consumption of trash food [1]. There are reports of children of East Asian or African ethnicity in Western nations having an increased risk of developing food allergies compared with Caucasian children. This suggests that adopting Westernized food habits could increase food allergies in African or Asian countries [2, 3].

The research about healthy, low-cost alternative products that can meet the enormous demands of a growing population involve legumes [4]. Legume crops represent a sustainability solution, serving as a fundamental source of high-quality alternative protein, reducing the emission of greenhouse gases, allowing the sequestration of carbon in soils, saving the CO_2 print thanks to the nitrogen fertilizer, it free high-quality organic matter that facilitate water retention and perform the soil nutrients circulation among others uses [5]. Despite their advantages, legumes contain proteins that can potentially cause food allergies. Several allergens from different legumes have been identified and characterized as proteins with potential allergic effects. These include lentil, pea, chickpea, soy, peanut, and lupine [6].

Clinically, the absence of sensibilization phase is a reliable indicator of the tolerance to an allergen. In this context, the presence of sensitization to a specific allergen protein has to be proven [7] both, the specific reactivity to a particular allergen protein and the cross-reactivity to other related allergens. The most frequent cross-reactivity process described clinically is that between lupin and peanut [8].

In Spain, consumption of legumes is common because they are an important part of the Mediterranean diet. It is estimated that consumption of legumes in Spain is 4.8 kg per year, with a greater percentage of children eating them as compared to adults. Legume consumption in Spain is greater in girls than in boys [9]. One study in Spain showed that food allergies were detected in 20.8% of children and 14% of adults. In the overall Spanish population, legumes were responsible of the 14.3% of the food allergies [10]. Another study of Spain's pediatric population found that 10% of children suffered from food allergies caused by lentil and 6.7% of children suffered from food allergies caused by peanuts. Lentil was found to be the most allergenic, causing 78% of reactions, followed by chickpea (72%) and peanut (33%) [9].

In Europe, legumes are the fifth-leading cause of food allergies [11]. A meta-analysis of studies conducted in Europe between January 2000 and September 2012 found that the percentage of the population with symptoms of food allergies plus specific immunoglobulin E (IgE) positivity activation to at least one food allergen was 3%–4.6% in children and 2.2%–2.66% in adults [12]. The same study concluded that the frequency of food allergy is greatest in northwestern European countries compared to southern European countries, which had the lowest prevalence. Some factors related to food allergies include environmental, genetic, and epigenetic factors that could suggest differences between global populations [13].

The general prevalence of food allergies is not clearly defined due to the lack of reliable data and the highly variable allergy patterns in different parts of the world. A selection of mixed developed country data (Allergy, Asthma & Immunology Research

2018) found that some allergies, like those to peanut, demonstrate heritability in Caucasian populations; skin immune responses shows differences between Asians and Caucasians. These types of studies have not yet been conducted in non-White populations, however, there exists some interest data showing that Black South African children present a significantly lower prevalence of peanut allergy compared to children of mixed-race origin (Black and Caucasian) by unknown factors [13].

One interesting fact about cross-reactivity is that it could be caused by proteins that come from species that are taxonomically distant. Examples of these antigens are panallergens, which are proteins conserved by evolution due to their important defense, structural, and storage functions [7]. If a person has an allergy to cow milk proteins, they are also probably allergic to goat milk proteins [14]. In the case of legumes, cross-reactivity to more than one legume is often found in children [9].

Overall, allergic features of allergen proteins could be attenuated by thermic proteolytic denaturalization due to the modification of the quaternary protein structure where superficial epitopes of these proteins' antigenic regions can still develop some allergenicity reactions. Despite this, there are studies that also show resistance to thermic, chemical, and proteolytic denaturalization, with is a common characteristic in legumes [15]. Some examples of resistance to denaturalization include allergen proteins like Cupins, very stable storage proteins that include legumins (11 S) and vicilins (7 S), both containing two common β-barrel structures in their globular domain. These appear to be a relevant stable structural motif, confirming resistance to denaturation and proteolysis [16]. Lipid transfer proteins (LTPs) have resistance to pepsin and to chemical digestion [17]; PR-proteins have thermostable structure [10] allowing them staying unalterable at physiological temperature. This stability plays an important role in allowing allergen active protein fragments to pass to the gastrointestinal tract, causing a food allergy.

There is a large public database of allergenic legume proteins with several isoforms. The commonly shared partial epitopes and their conservation in the same family of proteins in different species could be helpful in designing possible strategies to prevent cross-reactivity.

The aim of this work is to carry out an exhaustive molecular and structural analysis of the most common allergenic legume proteins through bioinformatic approaches.

2. Materials and methods

2.1 Search of legume proteins sequences

We used the Allergome and UniProt databases to search for allergenic legume proteins for this study. The proteins chosen are characterized by having complete sequences and being in mature form. The search was carried out on the available species of lentil, pea, chickpea, soybean, and lupine (**Table 1A-E**).

2.2 Alignment of sequences

The complete and mature sequences of lentil (*Len c 3*, *Len c 3.0101*, and *Len c aglutinin*), chickpea (*Cic a 1, Cic a 3, Cic a 4, Cic a 6*), pea (*Pis s 2* (7 s vicilin), *Pis s 3* (LTP), *Pis s 3.0101*(LTP), *Pis s 6* (PR-protein, Pis S aglutin, Pis s albumin)), lupine (*Lup a 1, Lup a alpha conglutin, Lup a delta conglutin, Lup a gamma conglutin, Lup a 4, Lup an 1, Lup an 1.0101, Lup an 3, Lup an 3.0101, Lup an alpha conglutin, Lup an delta*

Species	Protein name	Protein type	UniProtKB
A.			
Soy allergen sequences			
Glycine max	Gly m 5	Profilin	C6T9L1 (C6T9L1_SOYBN)
	Gly m 5.0301	Profilin	P25974 (GLCB1_SOYBN
	Gly m 8	2 s albumin	C6SYA7 (C6SYA7_SOYBN)
	Gly m 8.0101	2 s albumin	P19594 (2SS_SOYBN)
B.			
Selected sequences of Lupinus			
Lupinus albus	Lup a 1	7 s vicilin	Q53HY0 (CONB1_LUPAL)
	Lup a alpha conglutin	11 s conglutin	Q53I54 (Q53I54_LUPAL)
	Lup a delta conglutin	2 s albumin	Q333K7 (Q333K7_LUPAL)
	Lup a gamma conglutin	Aspartic protease	Q9FEX1 (CONG2_LUPAL)
	Lup a 4	PR-protein	O24010 (O24010_LUPAL)
Lupinus angustifolius	Lup an 1	7 s vicilin	B0YJF8 (B0YJF8_LUPAN)
	Lup an 3	LTP	A0A1J7GK90 (A0A1J7GK90_LUPAN)
	Lup an 3.0101	LTP	A0A4P1RWD8 (A0A4P1RWD8_LUPAN)
	Lup an alpha conglutin	11 s globulin	F5B8V6 (CONA1_LUPAN)
	Lup an delta conglutin	2 s albumin	F5B8W8 (COND1_LUPAN)
	Lup an gamma conglutin	Aspartic protease	Q42369 (CONG1_LUPAN)
Lupinus luteus	Lup l 4	PR- protein	P52778 (L18A_LUPLU)
C.			
Selected sequences of Pea			
Pisum sativum	Pis s 2	7 s vicilin	P13915 (CVCA_PEA)
	Pis s 3	LTP	A0A158V755 (NLTP2_PEA)
	Pis s 6	PR-protein	P13239 (DRR1_PEA)
	Pis s agglutinin	Agglutinin	B5A8N6 (B5A8N6_PEA)
	Pis s albumin	Albumin	P08688 (ALB2_PEA)
D.			
Selected sequences of Chickpea			
Cicer arietinum	Cic a 1	7 s vicilin	Q304D4 (Q304D4_CICAR)
	Cic a 3	LTP	O23758 (NLTP_CICAR)
	Cic a 4	PR-protein	Q39450 (Q39450_CICAR)
	Cic a 6	11 s globulin	Q9SMJ4 (LEG_CICAR)
E.			
Selected sequences of Peanut			
Arachis hypogaea	Ara h 1	7 s vicilin	B3IXL2 (B3IXL2_ARAHY)
	Ara h 1.0101	7 s vicilin	P43238 (ALL12_ARAHY)
	Ara h 2.0101	2 s albumin	Q6PSU2–2 (CONG7_ARAHY)

Species	Protein name	Protein type	UniProtKB
	Ara h 2.0201	2 s albumin	Q6PSU2–3 (CONG7_ARAHY)
	Ara h 3	11 s globulin	A1DZF0 (A1DZF0_ARAHY)
	Ara h 3.0201	11 s globulin	Q9SQH7 (Q9SQH7_ARAHY)
	Ara h agglutinin	Agglutinin	P02872 (LECG_ARAHY)
	Ara h 5	Profilin	D3K177 (D3K177_ARAHY)
	Ara h 5.0101	Profilin	Q9SQI9 (PROF_ARAHY)
	Ara h 6	2 s albumin	A1DZE9 (A1DZE9_ARAHY)
	Ara h 6.0101	2S albumin	Q647G9 (CONG_ARAHY)
	Ara h 7.0101	2 s albumin	Q9SQH1 (Q9SQH1_ARAHY9
	Ara h 7.0201	2 s albumin	B4XID4 (B4XID4_ARAHY)
	Ara h 7.0301	2 s albumin	Q647G8 (Q647G8_ARAHY)
	Ara h 8	PR- 10 protein	B1PYZ4 (B1PYZ4_ARAHY)
	Ara h 8.0101	PR-10 protein	Q6VT83 (Q6VT83_ARAHY)
	Ara h 8.0201	PR- 10 protein	B0YIU5 (B0YIU5_ARAHY)
	Ara h 9.0101	9 k-LPT	B6CEX8 (B6CEX8_ARAHY)
	Ara h 10.0101	16kD protein	Q647G5 (OL101_ARAHY)
	Ara h 11.0101	14KD oleosin	Q45W87 (OL111_ARAHY)
	Ara h 11.0102	14kD oleosin	Q45W86 (OL112_ARAHY)
	Ara h 13.0102	Defensine	C0HJZ1 (DEF3_ARAHY)
	Ara h 14.0101	17.5kD oleosin	Q9AXI1 (OL141_ARAHY)
	Ara h 14.0102	17kD oleosin	Q9AXI0 (OL142_ARAHY)
	Ara h 14.0103	17kD oleosin	Q6J1J8 (OL143_ARAHY)
	Ara h 15.0101	17kD oleosin	Q647G3 (OLE15_ARAHY)
	Ara h 16	7 k LPT	A0A445DA28 (A0A445DA28_ARAHY)
	Ara h 17	11 k LTP	A0A445AL51 (A0A445AL51_ARAHY)
Arachis duranensis	Ara d 2	2 s albumin	A5Z1Q8 (A5Z1Q8_ARADU)
	Ara d 6	2 s albumin	A5Z1Q5 (A5Z1Q5_ARADU)
Arachis ipaensis	Ara i 2	2 s albumin	A5Z1Q9 (A5Z1Q9_ARAIP)
	Ara i 6	2 s albumin	A5Z1Q6 (A5Z1Q6_ARAIP)
F.			
Selected sequences of Lentil			
Lens culinaris	Len c 3	LTP	A0AT28 (NLTP1_LENCU)
	Len c 3.0101	LTP	A0AT29 (NLTP2_LENCU)
	Len c agglutinin	Agglutinin	P02870 (LEC_LENCU)

Table includes the species name, the common name of the allergen, the type of protein according to its biological nature/function, and the UniProt entry name (UniProtKB). All sequences were used for alignment, T-cell epitope search, and IgE analysis. Sequences from all lupin and soybean species were used for the post-translational modification search tasks (A and B). For secondary and tertiary structure assessment, only the sequences of interest were used: G. max *(Gly m 5, Gly m 5.0301, Gly m 8, and Gly m 8.0101);* L. albus *(Lup a 1 and Lup a alpha conglutin);* Lupinus angustifolius *(Lup an alpha e);* P. sativum *(Pis s albumin);* C. arietinum *(Cic a 6); and* A. hypogaea *(Ara h 5.0101).*

Table 1.
Summary of the sequences used in successive studies.

conglutin, Lup an gamma conglutin, Lup l 4), and peanut (*Ara d 2, Ara d 6, Ara h 1, Ara h 1.0101, Ara h 2, Ara h 2.0101, Ara h 2.0201, Ara h 2.0202, Ara h 3, Ara h 3.0201, Ara h agglutin, Ara h 5, Ara h 5.0101, Ara h 6, Ara h 6.0101, Ara h 7.0101, Ara h 7.0102, Ara h 7.0301, Ara h 8, Ara h 8.0101, Ara h 8.0201, Ara h 9.0101, Ara h 10.0101, Ara h 11.0101, Ara h 11.0102, Ara h 13.0102, Ara h 14.0101, Ara h 14.0102, Ara h 14.0103, Ara h 15.0101, Ara h 16, Ara h 17)* were aligned by pairs against soybean allergens (*Gly m 5, Gly m 5.0301, Gly m 8, Gly m 8.0101*) extracting the identity percentage and comparing the possible differences in the amino acid nature of the protein sequences (positive charge, negative charge, and polarity) of the allergens listed above.

2.3 Functional domain analysis

We used the program Pfam v34.0 (http://pfam.xfam.org/) to identify the possible domains present in the isoforms of legume proteins.

2.4 Post-translational modification site prediction

We used the MusiteDeep deep learning framework (https://github.com/duolinwang/MusiteDeep_web) to search for the presence of possible post-translational modifications and identify how they affect the potential allergenicity of the study proteins [18]. The prediction models used are phosphorylation (Y, S, T); N-linked glycosylation (N); O-linked glycosylation (S, T); ubiquitination; N6-acetyllysine (K); Methylarginine (R); Methyllysine (K); Hydroxyproline (P) and Hydroxylysine (K) with a threshold value of 0.8.

S-nitrosylations and T-nitrations were also studied via the iSNO-AAPair tool (Y. Xu et al., 2013), which was used to predict cysteine S-nitrosylation sites (http://app.aporc.org/iSNO-AAPair) with a threshold value greater than 0.8. The GPS-YNO2 tool (Liu et al., 2011) was used to predict tyrosine nitration sites (http://yno2.biocuckoo.org).

2.5 Secondary structure assessment

Secondary structure was assessed using PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/). Sequence alignment was performed with CLUSTALW (https://www.genome.jp/tools-bin/clustalw), which was visualized with the BioEdit program, and in which the consensus secondary structure was annotated.

2.6 Modeling of three-dimensional structure

The three-dimensional structures of olive ALDH proteins were modeled using the Phyre2 web program (http://www.sbg.bio.ic.ac.uk/phyre2), which is based on Markov algorithms to generate alignments of the problem protein sequences with proteins with experimentally obtained protein crystallographic models (PDB).

2.7 Identification of IgE-binding epitopes

We used the AlgPred server (www.imtech.res.in/raghava/algpred/submission.html), which creates arrays using sequences from known allergens, to identify IgE-binding epitopes and to determine potential allergenicity of proteins based on of their amino acid and dipeptide composition.

2.8 Identification of T cell binding epitopes

We used the ProPred program (Singh et al., 2011) (http://webs.iiitd.edu.in/raghava/propred/) to analyze the protein sequences of legumes in the study. The analysis was performed with a 2% threshold for the most common human HLA-DR alleles among the Caucasian population: [DRB1*0101 (DR1), DRB1*0301 (DR3), DRB1*0401 (DR4), DRB1*0701 (DR7), DRB1*0801 (DR8), DRB1*1101 (DR5), and DRB1*1501 (DR2)].

3. Results and discussion

3.1 Sequences obtained from the Allergome database

We used the Allergome database to retrieve the available sequences of complete proteins of legumes, following the link to UniProt. The legumes included in this study are lentil, lupin, pea, chickpea, and peanut. Only two major allergens (*Gly m 5* and *Gly m 8*) with their available isoforms were extracted from soybean and used as reference to carry out the alignments and further analyses.

The reference proteins, soybean major allergens *Gly m 5* and *Gly m 8* with their isoforms, correspond to profilin, 7 s globulins, and albumin 2 s protein families. The allergen *Gly m 8* is considered to have the highest sensitivity [19], specificity, and reproducibility [20] to clinical reaction to soybean in atopic patients. The combination of *Gly m 5* and *Gly m 8* was suggested as one of the best ways to perform the estimation of the sensitization level and to improve the diagnosis of soybean allergy in children [21]. Thus, in the case of high similarity between the sequences of these soy allergens and the allergens of the other legumes included in this study, the diagnosis of possible cross-reactions between them could be facilitated.

3.2 Alignment of allergen protein sequences

Sequence alignments were performed to compare the common and differential features between allergen proteins and legumes. Overall, and according to the CODEX Alimentarius Commission in 2003, only proteins with a percentage of identity greater than 50% by local alignment (BLAST) are at risk of allergy or cross-reactivity [22]. Therefore, results obtained from protein–protein alignment beforehand do not show values high enough to make a prediction of possible cross-reactivity between soybean proteins and the rest of the legumes (**Table 2**).

The highest percentage of identity was the result of the alignment between the *Gly m 5* proteins and the *Gly m 5.0301* isoform (**Table 3**) with the *Lup a 1* protein with values of 48.41% and 48.72%, respectively (**Table 2D**). However, these percentages do not exceed the minimum alignment percentage recommended as guidance. Despite this, there are reported cases of cross-reactivity between other proteins with which there is a percentage lower than the standard minimum value considered for cross-reactivity and lower than that which occurs between these proteins, as in the case of *Gly m 8* and *Ara h 2* [23], with an identity percentage of 31.46% (**Table 2F**).

The multiple alignment analysis between *Gly m 5* and the isoform *Gly m 5.0301* with the *Lup a 1* protein obtained a percentage of common identity of 35.80% with 207 identical positions (Image 1).

Glycine max	Arachis duranensis		*Lens culinaris*		
Protein name	**Ara d 2**	**Ara d 6**	**Len c 3**	**Len c 3.0101**	**Len c agglutinin**
A.					
Percentages of Amino Acid Sequence Identity by Alignment of Peanut and Lentil Species against Reference Soybean Sequences					
Gly m 5	8428	6067	5239	5157	11.803
Gly m 5.0301	9009	5909	4556	5817	11.349
Gly m 8	32.738	29.94	9942	10.465	9375
Gly m 8.0101	33.333	29.94	9942	9884	9278

Glycine max	*Lupinus angustifolius*						
Protein name	**Lup an 1**	**Lup an 10,101**	**Lup an 3**	**Lup an 30,101**	**Lup an alpha conglutin**	**Lup an delta conglutin**	**Lup an gamma conglutin**
B.							
Percentages of Amino Acid Sequence Identity by Alignment of Lupin Species against Reference Soybean Sequences							
Gly m 5	24.463	39.739	5843	4	17.304	8444	15.028
Gly m 5.0301	24.463	39.739	4.31	4	17.304	8444	14.657
Gly m 8	8114	6209	11.561	11.243	6616	35.62	5298
Gly m 8.0101	7877	6209	12.069	11.765	6616	36.25	5066

Glycine max	*Cicer arietinum*				Arachis ipaensis	
Protein name	**Cic a 1**	**Cic a 3**	**Cic a 4**	**Cic a 6**	**Ara i 2**	**Ara i 6**
C.						
Percentages of Amino Acid Sequence Identity by Alignment of Chickpea and Peanut Species against Reference Soybean Sequences						
Gly m 5	36.759	6378	8	13.587	8753	6292
Gly m 5.0301	37.575	6378	7556	8	8.85	6136
Gly m 8	7143	10.526	5021	7585	31.461	29.94
Gly m 8.0101	6513	10.526	5021	7585	31.461	30.539

Glycine max	*Lupinus albus*					
Protein name	**Lup a 1**	**Lup a 4**	**Lup a alpha conglutin**	**Lup a delta conglutin**	**Lup a gamma conglutin**	**Lup l 4**
D.						
Percentages of Amino Acid Sequence Identity by Alignment of Lupin Species against Reference Soybean Sequences						
Gly m 5	48.417	6.25	16.637	8036	13.645	6798
Gly m 5.0301	48.717	6.25	16.637	8259	14.098	7456
Gly m 8	5151	10.698	6501	35.625	4425	13.115
Gly m 8.0101	5009	10	6.18	36.25	4435	13.115

Glycine max	***Pisum sativum***					
Protein name	**Pis s 2**	**Pis s 3**	**Pis s 3.0101**	**Pis s 6**	**Pis S agglutin**	**Pis s albumin**
E.						
Percentages of Amino Acid Sequence Identity by Alignment of Pea Species against Reference Soybean Sequences						
Gly m 5	41.638	5467	5882	6798	9362	6798
Gly m 5.0301	41.638	5145	5369	6798	11.429	10.444
Gly m 8	5759	11.765	10.588	13.402	11.273	8.98
Gly m 8.0101	5.41	11.176	10.588	13.402	10.204	9388

Glycine max	***Arachis hypogaea***							
Protein name	**Ara h 1**	**Ara h 1.0101**	**Ara h 2**	**Ara h 2.0101**	**Ara h 2.0201**	**Ara h 2.0202**	**Ara h 3**	**Ara h 3.0201**
F.								
Percentages of Amino Acid Sequence Identity by Alignment of Peanut Species against Reference Soybean Sequences								
Gly m 5	36.585	35.726	8753	8811	8874	9031	15.412	14.685
Gly m 5.0301	36.748	35.885	8.85	9009	9292	9234	15.762	14.86
Gly m 8	5769	8307	31.461	32.738	34.818	33.133	5.41	6015
Gly m 8.0101	7329	7668	31.461	33.333	31.818	33.735	4.57	5636

Glycine max	***Arachis hypogaea***							
Protein name	**Ara h agglutinin**	**Ara h 5**	**Ara h 5.0101**	**Ara h 6**	**Ara h 6.0101**	**Ara h 7.0101**	**Ara h 7.0102**	**Ara h 7.0301**
G.								
Percentages of Amino Acid Sequence Identity by Alignment of Peanut Species against Reference Soybean Sequences								
Gly m 5	13.816	6349	6136	6606	6292	7982	7062	6292
Gly m 5.0301	13.717	4904	5.33	6951	6136	8296	6834	9131
Gly m 8	8571	10.734	6015	28.144	29.94	23.497	30.337	22,286
Gly m 8.0101	8571	10.674	9091	28.144	30.539	23.497	30.899	22.857

Glycine max	***Arachis hypogaea***							
Protein name	**Ara h 8**	**Ara h 8.0101**	**Ara h 8.0201**	**Ara h 9.0101**	**Ara h 10.0101**	**Ara h 11.0101**	**Ara h 11.0102**	**Ara h 13.0102**
H.								
Percentages of Amino Acid Sequence Identity by Alignment of Peanut Species aAgainst Reference Soybean Sequences								
Gly m 5	6181	7761	6.92	3596	7761	6982	7207	3139
Gly m 5.0301	6935	7539	6.92	3.82	6828	6982	7207	3139

Glycine max	*Arachis hypogaea*							
Protein name	**Ara h 8**	**Ara h 8.0101**	**Ara h 8.0201**	**Ara h 9.0101**	**Ara h 10.0101**	**Ara h 11.0101**	**Ara h 11.0102**	**Ara h 13.0102**
Gly m 8	11.429	10.233	11.64	10.405	6478	6.14	6.14	9877
Gly m 8.0101	11.792	11.64	11.64	10.405	6883	6.14	6.14	9259

Glycine max	*Arachis hypogaea*					
Protein name	**Ara h 14.0101**	**Ara h 14.0102**	**Ara h 14.0103**	**Ara h 15.0101**	**Ara h 16**	**Ara h 17**
I.						
Percentages of Amino Acid Sequence Identity by Alignment of Peanut Species against Reference Soybean Sequences						
Gly m 5	8744	7848	8296	7221	4698	3905
Gly m 5.0301	8744	7848	8296	7221	4698	4121
Gly m 8	5785	5859	5785	5.6	11.111	11.243
Gly m 8.0101	5372	5859	5785	5.6	11.31	11.243

Degree of identity resulting from the alignment of amino acid sequences. These have been obtained by alignment between soybean proteins, used as reference, against different legume species (lentil, chickpea, pea, lupine, and peanut) including major allergens and isoforms.

Table 2.
Percentages of amino acid sequence identity by alignment of different legume species against reference soybean sequences.

Alignment Frequency Calculations			
Average of the difference of the frequencies between the different isoforms of soybean proteins with the alignment of the different proteins of legume species.			
Gly m 5/Gly m 5.0301	0,599 (over all)	**values > 3%**	5587 (Cic a 6) 3646 (Pis s albumin)
Gly m 8/Gly m 8.0101	0,468 (over all)	**values > 3%**	3076 (Ara h 5.0101)
Max identity values obtained by sequences alignment			
Greater value	48,717 (over all)		
	Gly m 5.0301 vs. Lup a 1		

Table 3.
Summary of the largest (greater than 3%) and smallest differences as a result of legume–soy protein alignment.

These data show that the percentage of identity of allergens must be kept in mind to compare allergens and to predict potential allergenicity and cross-reactivity, since not only do sequential epitopes have to be taken into account for that purpose, but also 3D and specific structural conformations of particular allergen proteins must be considered.

Using the information obtained by alignment, some of the proteins in the comparative analysis with soybean could be of interest at the molecular allergy level, such as Lup a delta conglutin and Lup an delta conglutin with percentages of identity with *Gly m 8* and *Gly m 8.0101* ranging from 35 to 36%. It also presents notable alignment

percentage differences with *Gly m 5* and *Gly m 5.0301* (**Table 2B, D**), with approximately 8% being the most notable difference in identity with respect to the other conglutins. Another candidate protein for analysis is *Lup a delta conglutin* with percentages of identity of 35.63% and 36.25% compared to *Gly 8* and its isoform *Gly m 8.0101*, respectively (**Table 2D**) and *Lup an delta conglutin* of 35.62% and 36.25%, respectively (**Table 2B**). The identity ratios are lower than the minimum value considered to establish cross-reactivity with soybean. However, with such similar percentages among conglutin sequences it is worthy to conduct a deeper analysis. Multiple alignment shows a high rate of conservation between lupin proteins from the different species of *L. albus* and *Lupinus angustifolia*. Comparison of gamma conglutin protein sequences of both species, soybean obtained a low identity percentage of 13–15% compared to *Gly m 5* and 4–5% compared to *Gly m 8* (**Table 2B, D**). Alignment between both conglutins showed an identity of 84.21%, with 128 identical positions and 12 similar positions (**Figure 1**), with an identity value high enough to consider cross-reactivity among them. Indeed, these sequences showed high conservation rate among lupin proteins from different species such as *L. albus* and *L. angustifolia*. The three-dimensional structure of these conglutins will be further analyzed in later sections (**Figure 2**).

Considering the identity percentages previously indicated, the Ara h 2 identity percentage of 31% at *Gly m 8* with demonstrated cross-reactivity and the 48% identity of *Lup a 1* with soybean, we found more cases of proteins with intermediate values. Such is the case of *Pis s 2* with *Gly m 5* and its isoform with an identity of 41.638% (**Table 2E**) and *Cic a 1* with 36.76% and 37.58% identity with *Gly m 5* and its isoform, respectively (**Table 2C**). On the other hand, the characterization of demonstrated cross-reactivity between soybean and peanut, as is the case of *Ara h 1* with *Gly m 5* and its isoform *Gly m 5.0301,* showed a 36.59% and 36.75% identity, respectively [24]. The rest of the alignments show percentages less than the described data of identity range and may be discarded from the depth in their CR study (**Table 2**).

Interestingly, the percentage of alignment identity between soybean isoforms was low, with values less than 1%, specifically, in the alignment of soybean major allergen *Gly m 5* and its isoform *Gly m 5.0301*. The sequences of these two allergens were compared to the rest of the legume proteins considered in this study. We obtained a different percentage of identity of 0.6%, as well as 0.47% when compared *Gly m 8* with *Gly m 8.0101* (**Table 3**). The largest differences were found between soybean isoforms and legumes; *Gly m 5/Gly m 5.0301* was 5.60% against chickpea protein *Cic a 6* (**Table 2C**); 3.65% against pea *Pis s albumin* (**Table 2E**) protein; and *Gly m 8 /Gly m 8.0101* 3.07% against peanut (*A. hypogaea*) protein *Ara h 5.0101* (**Table 2G**). **Table 3** summarizes this data.

The existence of differences between isoforms of other legume species of the same allergen protein family could open the way for new studies finding significant differences in multiple cross-reactivity candidacy. For example, such as the case of *Lup an 1* and *Lup an 1. 0101* with identity differences exceeding 13% in alignment with *Gly m 5*, and ranging between 24.46% and 39.74%, respectively (**Table 2B**). These differences make *Lup an 1* an unsuitable candidate for cross-reactivity, whereas its isoform *Lup an 1.0101* could be a candidate for cross-reactivity with soybean.

3.3 Post-translational modification analysis

Post-translational modifications affecting the allergen protein sequences have been defined and involved in processes like alcohol or tiol addition (glycosidations), methyl

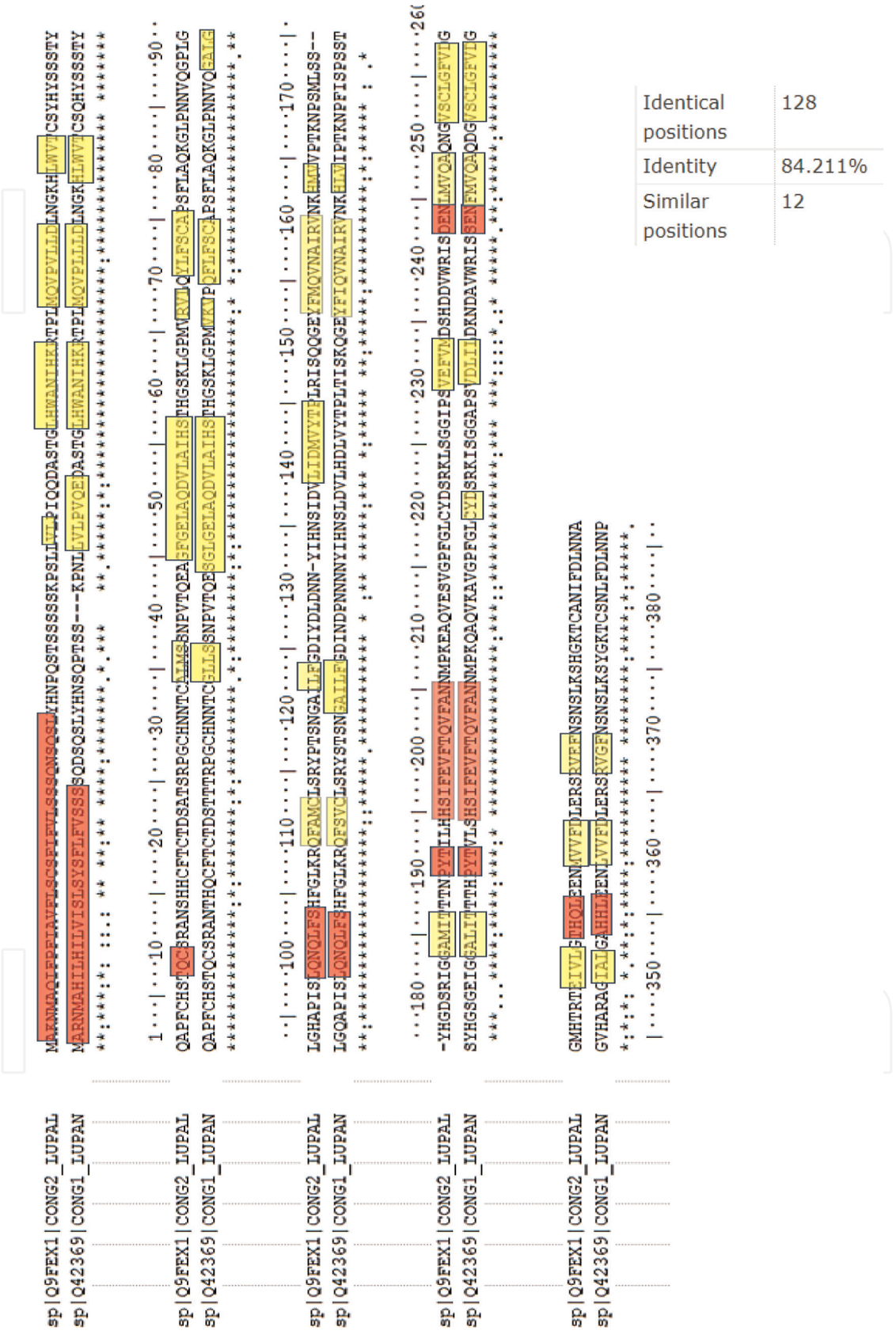

Figure 1.
*2D structure of allergen proteins. Multiple alignment of the major Lup a gamma conglutin (*Lupinus albus*) against Lup an gamma conglutin (*Lupinus angustifoluis*) with the secondary sequence represented in yellow by coil zones and in red by helix zones. In addition to the percentage of joint identity, number of identical amino acid positions and number of amino acid have similar physicochemical nature.*

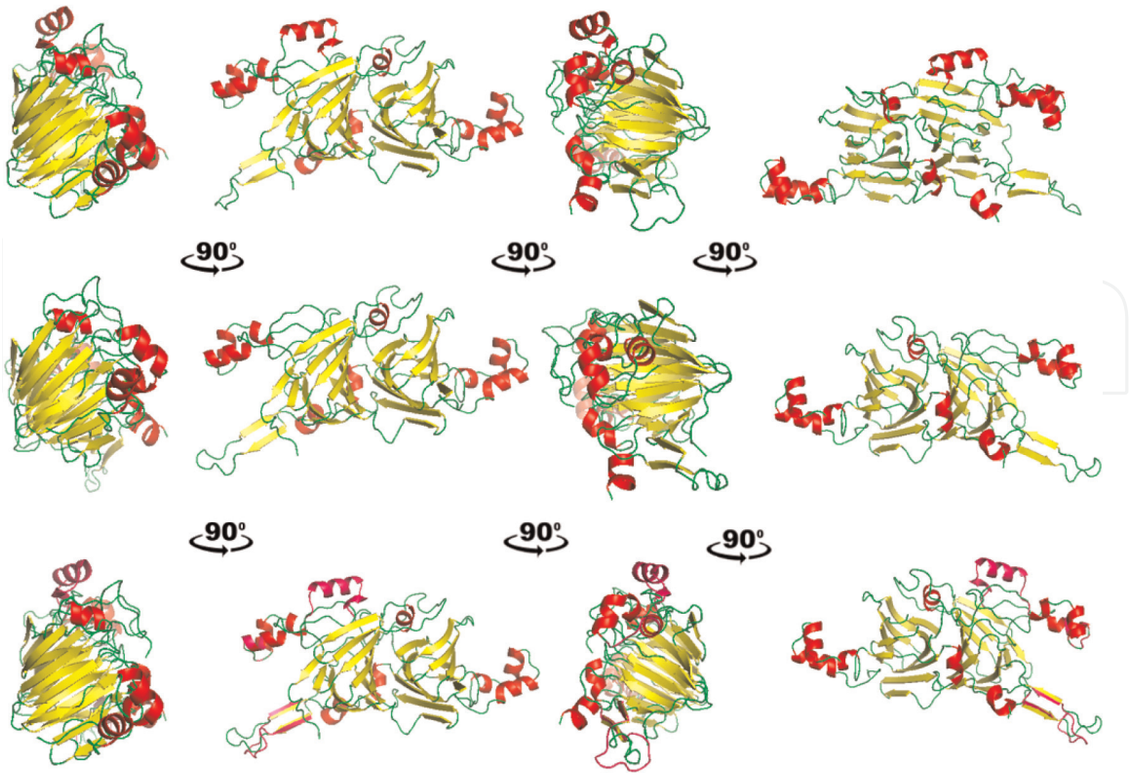

Figure 2.
Three-dimensional structural analysis of seed allergen proteins. Figures of first row corresponding to the 3D structures of the Lup a gamma conglutin protein; second row represent different views of Lup an gamma conglutin; and third raw are the figures of the consensus sequence with depicted match regions in pink color over the consensus figure (last row). Red color highlights the alpha-helix and yellow color the beta-strand.

groups (methylations), phosphates (phosphorylations), carboxyl groups (carboxylations), nitro groups (T-nitrations), or nitroxil groups (S- nitrosylations).

These types of modifications may induce rearrangements in structure, which could indirectly affect lineal and/or conformational epitopes' influence pm molecular allergy, limiting or favoring immunological recognition as well as generating antigenic diversity [25]. It is interesting to analyze location of where these modifications may occur and the type of modification together with the influence of these modifications in the 2D structural elements.

Phosphorylation is considered a factor of change of molecular pH dynamics [26], generating important alterations in the biophysics of the protein [27]. It has been observed sites of phosphorylation in most of the proteins examined: *Gly 5, Gly 8* and their isoforms; *Lup a 1, Lup a* alpha and delta conglutins (*L. albus*); *Lup an 1* and its isoform *Lup an 1.0101, Lup an alpha, Lup an delta* and *Lup an gamma (L. angustifolius)*. In the sequences of *Lup l 4 (L. luteus)* and *Cic a 6 (C. arietinum)* are also abundant modifications as glycosidations which potential importance in the allergenicity behavior of these proteins. In this regard, it has been demonstrated in some cases the increasing immunogenicity [28] for *Gly 5* and *Gly 8*; *Lup a 1, Lup a 4*, *Lup a alpha, delta*, and *gamma conglutins*; *Lup an 1* and it isoform *Lup an 1.0101*, *Lup an alpha* and *gamma conglutins*; *Lup l 4* and *Cic a 6* (**Table 4**).

Methylations are quite less abundant modifications. It is observed that their deficiency generates serious alterations in the functioning of proteins, thus having important implications on their three-dimensional structuring as carboxylation [29]. Only two methylation sites were found: one on *Lup a alpha conglutin* and one on *Lup an alpha conglutin* (**Table 4B**). Carboxylations were found on the *Gly m 8.0101* isoform;

Lup a alpha, delta, and *gamma conglutins*; *Lup an 1* and its isoform *Lup an 1.0101;* and *Lup an 3* and *Lup an alpha conglutin* (**Table 4A, B**).

Nitrosylation and nitrations generate strong covalent bonds in the protein structure [30, 31]. Nitrations were found on *Lup a 1*, *Lup a 4*, and *Lup a alpha conglutin*; *Lup a gamma conglutin*, *Lup an 1,* and *Lup an 1.0101*; *Lup an 3.0101*, *Lup an alpha* and *gamma conglutin*; *Lup l 4*; *Cic a 6* and *Ara h 5.0101*. Nitrosylations in comparison were less abundant, found in *Lup a alpha conglutin*; *Lup an 3* and its isoform *Lup an 3.0101*, and *Lup an alpha, delta,* and *gamma conglutins* (**Table 4**).

Post-translational modifications on T-cell epitopes have been found in *Gly m 5*.0301 isoform, a glycosidation at position 351, and a nitration at 172; *Lup a alpha conglutin* presents three methylation sites at positions 199, 448, and 497; *Lup a delta conglutin* contains a glycosidation site at position 76; a nitrosylation site at position 13 was found in *Lup an 3*, while in its isoform a nitration at position 104 and a nitrosylation at position 112 are highlighted; *Lup an delta* conglutin presents a candidate phosphorylation site at position 76 and Cic a 6 a nitrosylation at 107. In other cases, IgE epitopes are affected, with the only case of *Lup a alpha* conglutin with a methylation site at position 102. **Table 5** presents a summary of this data.

The direct implications of these post-translational modifications may be directly linked to the effects on the variation of the structure of these regions, generating differential epitopes recognition and consequently the allergen response.

Analyzing the location and type of modifications could help to elucidate the relationship of protein structure epitope distribution to the allergen potential of the protein, however, it will not be confirmed whether the different modifications would accentuate or lessen the allergenic impact until a clinical review of the process is carried out. The possibility of inducing post-translational modifications on plant proteins as a therapeutic tool is being examined [27].

3.4 Secondary structure analysis

The combined analysis of secondary structure with multiple alignments allows a direct sequence–structure–functional comparation between different allergen proteins. An interesting analysis has been made to identify the areas of allergens with shared mutual domains as part of structural domains with important implications for cross-reactivity potential.

The *Gly m 5, Gly m 5.0301*, and *Lup a 1* secondary structure comparison showed that in sequences of these proteins (**Table 2A**), the percentage of identity with *Lup a 1* was the highest compared to the rest of the alignments performed (**Table 3**). However, the percentage was not potentially enough to induce cross-reactivity. Comparative analysis between the secondary structure predictions of these proteins shows strong similarities in the distribution of α-helix and β-strand over middle regions of the proteins (amino acids 20–430) (**Figure 3**), giving an additional perspective of the possible regions with potential cross-reactivity in addition to the information provided by the alignments.

The three allergen proteins include Cupin superfamily domains with a wide variety of representative enzymes, but notably contains the non-enzymatic seed storage proteins [32]. Functional domains that could be candidates to potentially undergo post-translational modifications for *Lup a 1* are one of the two barrel domains with anti-parallel b-sheets. The first one is a Cupin_1.1 (**Table 6A**), a candidate for glycosidation (**Table 4B**). Similarly, in the case of *Gly m 5* and its isoform *Gly m 5.0301*, in both proteins where also present these modifications in their globular

Allergen	Post-translational modifications					
	Phosphorylation	**Glycosylation**	**Pyrrolidone carboxylic acid**	**Methylation**	**Nitration**	**Nitrosylation**
A.						
Post-translational modifications predicted over soybean: Glycine max (Gly m)						
Gly m 5	232; 234; 235	351	—	—	158;172	—
Gly m 5.0301	232; 234; 235	351	—	—	158; 172	—
Gly m 8	155; 156	120	—	—	—	14
Gly m 8.0101	155; 156	120	25	—	—	14
B.						
Post-translational Modifications Predicted Over Lupinus: Lupinus albus (Lup a), L. angustifolius (Lup an), and L. luteus (Lup l)						
Lup a 1	71; 79; 104	444	—	—	269;316	—
Lup a 4	—	13; 82	—	—	157; 269; 316	—
Lup a alpha conglutin	347	403	29	102	199; 448; 497	36; 334
Lup a delta conglutin	75;76	73; 108	27	—	—	—
Lup a gama conglutin	—	133	28	—	261	—
Lup an 1	80;82;85	152; 434	126; 158	—	340	—
Lup an 1.0101	80;82;85; 469; 488	434; 519	126; 158	—	340; 488	—
Lup an 3	—	—	23	—	—	13; 27
Lup an 3.0101	—	—	—	—	104	28; 112
Lup an alpha conglutin	247; 259; 341	397; 439	24	97	84; 442; 491	31
Lup an delta conglutin	76; 77; 80;83	—	—	—	—	42
Lup an gamma conglutin	357	130	—	—	259	350; 391; 440
Lup l 4	112	78; 82	—	—	100; 156	—
C.						
Post-translational Modifications Predicted Over Chickpea: Cicer arietinum (Cic a) and Peanut: Arachis hypogaea (Ara h)						
Cic a 6	139; 195; 207; 225; 271	1; 220	—	—	443	64; 107
Ara h 5.0101	—	—	—	—	6; 125	115

Specific amino acids affected by each type of post-translational modification on the different legume proteins: phosphorylation, glycosylation, carboxylation (pyrrolidone carboxylic acid), methylation, nitrosylation, and nitration sites. The (—) symbol means no results.

Table 4.
Post-translational modifications predicted over legumes.

Allergen name	Post-translational Modifications				
	Phosphorylation	**Glycosylation**	**Methylation**	**Nitration**	**Nitrosylation**
A.					
T-cell epitopes from allergens affected by post-translational modifications					
Gly m 5.0301		FVVNATSNL (351)		YLQGFDHNI (172)	
Lup a alpha conglutin			FGPLRRCN (199)		
			YVLNGSAWF (448)		
			YVAFKTNDI (497)		
Lup a delta conglutin		LVAALVLVV (76)			
Lup an 3					VLICMVVVS (13)
Lup an 3.0101				YKISTSTNC (104)	YKISTSTNC (112)
Lup an delta conglutin	LVVHTSASR (76)				
Cic a 6					FGMVFPGCV (107)
B.					
IgE epitopes from allergens affected by post-translational modifications					
Lup a alpha conglutin			IETWNPNNQEFECAG (102)		

This table summarizes the T-cell and IgE epitopes directly affected by the main post-translational modifications indicating the amino acid number affected.

Table 5.
T-cell and IgE epitopes from allergens affected by post-translational modifications.

domain (antiparallel β-barrels) (**Table 6A**), which is a candidate to undergo glycosylation (**Table 4A**). In three cases, modifications by glycosidation of one of their functional domains is a shared functional and allergenic feature.

Lup a gamma conglutin and *Lup an gamma conglutin* were analyzed. Although they belong to different species of lupin, they showed few differences in alignment and their comparison with soybean proteins of reference (**Table 2B, D**). The identity percentage among them is greater than 50%. These allergen proteins could be considered to exhibit CR, due to sequence identity but also to similarities of their secondary structure (**Figure 1**).

Regarding the predictions of post-translational modifications of these proteins relevant to 2D structural domains, it was found that *Lup a gamma conglutin* can be modified by a potential glycosidation (**Table 4B**). This modification is located in the

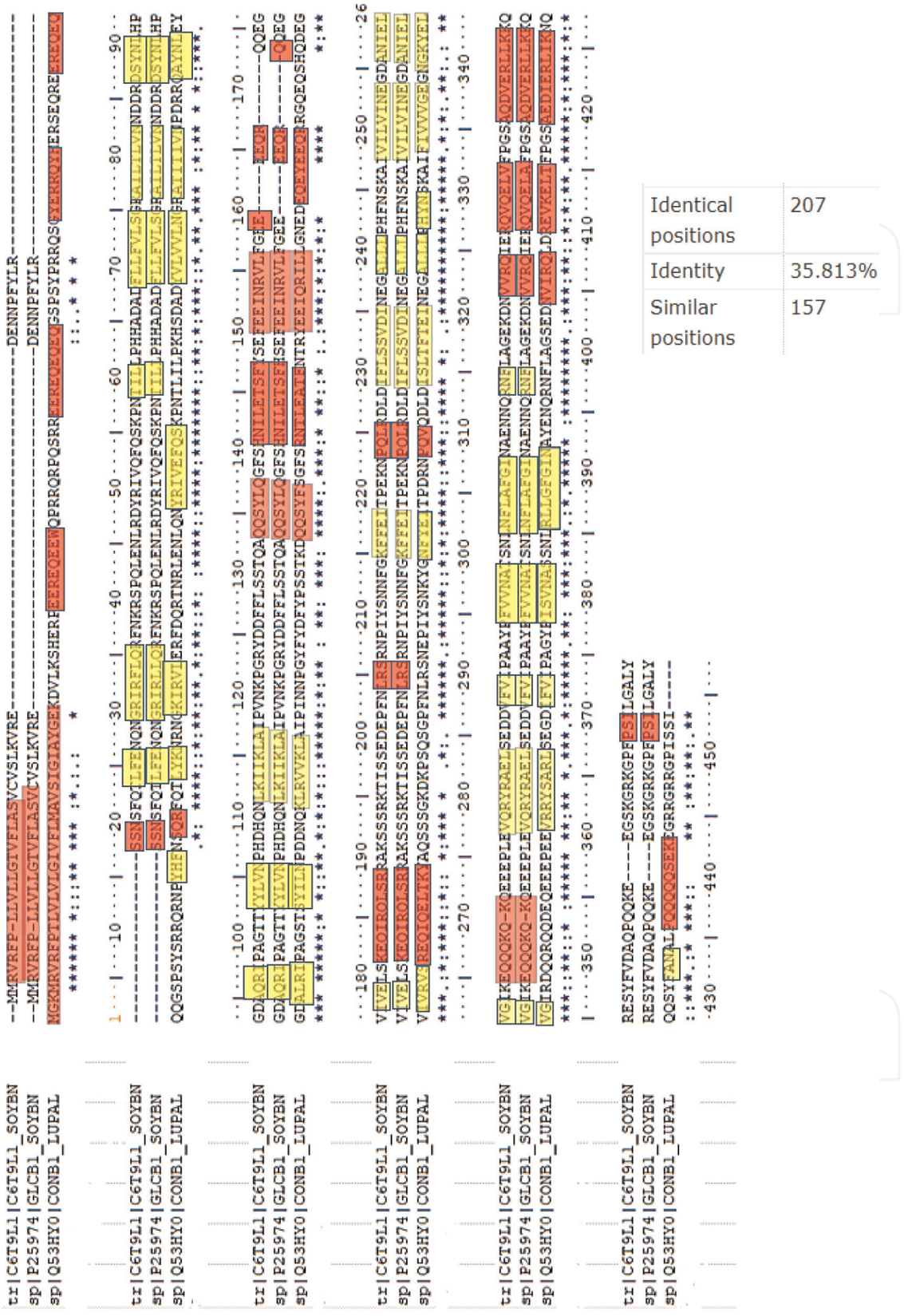

Figure 3.
*2D structure of allergen proteins. Multiple alignment of the major allergen Gly m 5, its isoform Gly m 5.0301 from (*Glycine max*) and Lup a 1 (*Lupinus albus*) together with the secondary sequence is represented in yellow by coiled-coil zones and in red by helix zones. In addition to the percentage of joint identity, number of identical amino acid positions and number of amino acid have similar physicochemical nature.*

Protein	Functional domain	Alignment amino acid range
A.		
Functional Domains Predicted Over Gly m 5, Gly m 5.0301 and Lup a 1		
Lup a 1	Cupin_1.1	332–486
	Cupin_1	137–227
Gly m 5	Cupin_1	240–389
	Cupin_2	86–144
Gly m 5.0301	Cupin_1	240–393
	Cupin_2	86–144
B.		
Functional Domains Predicted Over Lup a gamma conglutin and Lup an gamma conglutin		
Lup a gamma conglutin	Xylanase inhibitor C-terminal	271–428
	Xylanase inhibitor N-terminal	66–240
Lup an gamma conglutin	Xylanase inhibitor C-terminal	269–429
	Xylanase inhibitor N-terminal	63–237

This table summarizes the protein domains of the different proteins in their different types, specifying the range of amino acids that occupy in alignment.

Table 6.
Functional domains predicted over legumes allergens.

region of the protein domain xylanase inhibitor C-terminal (**Table 6B**). *Lup an gamma conglutin* has two possible domains affected by post-translational modifications: a phosphorylation and two nitrosylations (**Table 4B**) that affect the region comprised in the C-terminal xylanase inhibitor domain (**Table 6B**) and two nitrosylations (**Table 4B**) over the same domain. It also presents a glycosidation (**Table 4B**) in the xylanase inhibitor N-terminal domain (**Table 6B**).

3.5 Three-dimensional structure analysis

Analysis of three-dimensional structure of proteins (**Figure 4**) provides insight into their sequence conformation and epitope arrangement. It also helps to determine the consequences of possible structural changes occurring between protein isoforms with minimal or large number of changes (**Table 2**) in their sequences [33].

Post-translational modifications over protein domains also may generate changes in their three-dimensional structure, affecting exposure epitopes and increasing or decreasing their allergenic potential.

Some candidates to examine the three-dimensional structure are *Gly m 5*, *Gly m 5.0101*, and *Lup a 1* that share common barrel domains with alternating folds between the α-helix and β-strand. These domains are in a special conformation, forming a solenoid in which the β-strand is arranged on the inside of the toroid and the α-helix is arranged on the outside in the same domain (**Figures 2** and **5**).

The structural differences observed in the consensus structure between the three structures indicate that in *Gly m 5.0301*, an element of the 2D structure corresponding to a β-strand structural connection is not present in the isoform *Gly m 5*. Neither is it present in *Lup a 1*, which is a specific and important structural feature that can make a

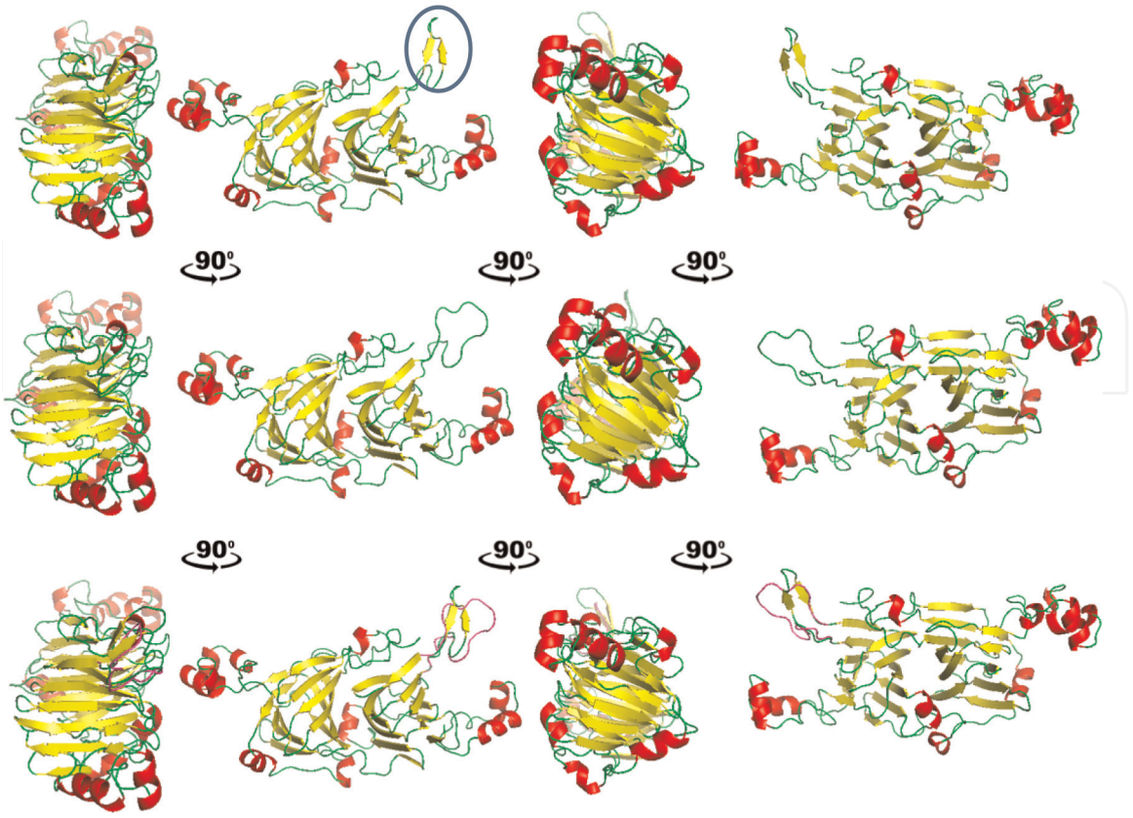

Figure 4.
3D structural analysis of seed allergen proteins. Three-dimensional structures of the Gly m 5.0301 *proteins are described, followed by* Gly m 5.0301 *and the change points between the two proteins marked in soft pink color in consensus figure (last row). Red denotes the alpha-helix and yellow denotes the beta-strand. T-epitope location is marked by a blue circle.*

specific conformational epitope (**Figures 4** and **5**). This structural change does not contain any epitope sequence. However, the change found is located between the *Cupin-1* domain of *Gly m 5* and its isoform, whereas this change in *Lup a 1* is located in the Cupin_1.1 domain (**Table 6A**).

Tridimensional structure comparison between Lup a gamma conglutin and Lup an gamma conglutin result on two principal differences observed between both conglutins, which is an α-helix in the gamma conglutin of *L. albus* that is not present in *L. angustifolius* (**Figure 2**). Regarding post-translational modification sites, in this loop there are no predicted modifications in this region encompassing the N-terminal xylanase inhibitor domain (**Table 6B**).

The 3D analysis was useful to determine other cases of interest previously mentioned, such as *Pis s 2* and *Cic a 1* in comparison with *Gly m 5* and its isoform that showed considerable identity ratios (**Table 2C, E**). *Lup an 1* and *Lup an 1.0101* showed large differences between their identity, and even more differences were found when compared to *Gly m 5*, which is somehow reflected in their 3D structures.

3.6 Identification and analysis of T-cell binding epitopes

An epitope is the portion of a macromolecule that is recognized by the immune system, specifically the sequence to which antibodies, B-cell receptors or T-cell receptors, can bind to initiate an immune response. Analysis of the epitopes shared for specific allergen proteins could be relevant to identify potential cross-reactivity.

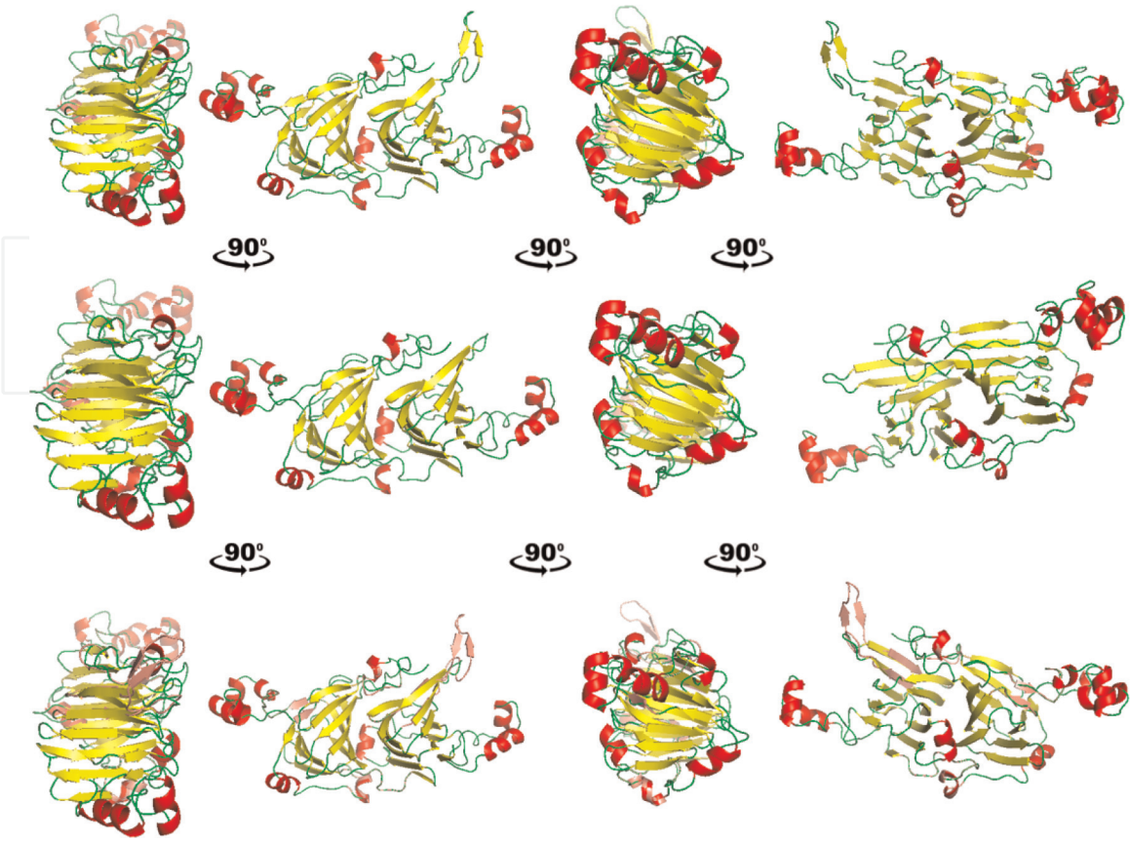

Figure 5.
3D structural analysis of seed allergen proteins. Three-dimensional structures of the Gly m 5 *proteins followed by* Lup a 1 *and representative changes between these two proteins marked in pink in the consensus figure (last row). Red denotes the alpha-helix and yellow denotes the beta-strand.The three-dimensional structure of the proteins* Gly m 5, Gly m 5.0301 *(*Glycine max*), and* Lup a 1 *(*L. albus*) showed a structure with large number of similarities, which is also reflected in the previous analysis of their secondary structure (**Figure 3**), with two barrel domains common in all of them.*

Presence of common T-cell epitopes among different legume species may support cross-reactivity processes; the greater the probability of occurrence, the larger the number of common epitopes.

The data obtained from the analysis of T-cell epitopes allows us to know which epitopes are shared among allergen proteins in the different legume species and to examine possible cases of cross-reactivity. Thus, in the case of soybean *G. max*, epitopes common to peanut, *A. hypogaea* species and chickpea *C. airietinum* species are described in **Table 7A**. It is remarkable that the soybean protein isoform *Gly m 5.0301* has an epitope in common with *Ara h 9.0101*, while the major allergen *Gly m 5* does not contain this epitope (**Table 7A**). This feature may be related to the cross-reactivity between specific sequences and these legume cultivars containing these specific proteins.

On the other hand, the different lupin species show that up to 18 T-cell epitopes are found commonly shared between *L. albus* and *L. angustifolius* (**Table 7B** part 1, 2, 3 and 4). Shared epitopes are also observed between *L. albus* and *A. hypogaea* (four epitopes) (**Table 7B** part 1, 2 and 4); *A. duranensis* (one epitope), *C. arietinum* (same number of epitopes) (**Table 7** part 1). Comparison with *L. angustifolius* showed three epitopes commonly shared with *A. hypogaea* (**Table 7B** parts 2, 3 and 4), and one epitope with *C. arietinum* and *L. culinaris* (**Table 7B** part 3).

Among these allergen proteins, there are also epitopes shared more than one time among more than two species. The same epitope is shared among the allergenic proteins: *Lup a 4* with *Ara h 8.0101* and *Cic a 4* (**Table 7B** part 1); *Lup an alpha conglutin*, *Lup an 3.0101, Ara h 3*, and *Ara h 3.0201* (**Table 7B** part 4). the most shared epitope was between *Lup an 3*, *Lup an 3.0101*, *Ara h 9.0101*, *Ara h 17*, *Cic a 3*, *Len c 3*, and *Len c 3.0101* (**Table 7B** part 3).

Prediction of secondary and tertiary structures allowed us to determine the spatial location of epitopes in proteins and to assess whether they may be affected in their spatial arrangement by post-translational modifications in protein domains over interest proteins.

Gly m 5, *Gly m 5.0301*, and *Lup a 1* analysis also showed that T-epitope regions founded over these proteins integrate part of the functional barrel domains of these proteins. In the case of *Gly m 5*, a single T-epitope (**Table 6A**) is located in the region of the structural domain between β-strands (**Figure 5**). This region is located into Gly m 5-barrel domain (Cupin_1) (**Table 6A**) in the amino acidic region located close to the site of glycosidation (**Table 5A**). This structural epitope is of special interest by its specificity, location, and potential specific allergenicity induced by this protein.

The T-cell epitopes analyzed on *L.* gamma conglutins resulted in the presence of two epitopes on the C-terminal xylanase and one on the N-terminal xylanase domain of *L. albus* (**Table 6B**, **Table 7B** part 1and 2) and one over N-terminal xylanase domain of *L. angustifolius* (**Table 6B** and 7 part 1). These are not directly or proximally affected by post-translational modifications, but they do affect the domains in which they are located.

Therefore, epitopic regions matched between *L. albus* and *L. angustifolius* conglutin, which are the most abundant compared to other epitopes (**Table 7B**). This supports the idea of conservation of protein structures and evidences the data found by simple comparative alignment.

3.7 Identification and analysis of IgE-binding epitopes

The IgE antibodies are produced by immune B cells, which in turn are stimulated by T cells responsible for recognizing the epitope in a sensitization step. To trigger the allergen inflammatory process, IgE antibodies stimulate the release of histamines. Thus, the recognition of these sequences allows for predicting the recognition capacity of IgE antibodies and whether they will potentially trigger the allergenic response (**Figure 6**).

The analysis of the allergenic nature of the protein based on amino acid and dipeptide analysis composition has been used for the assessment of the above proteins. It is noticeable that the 30cases with clinically confirmed allergenic epitopes are predicted by their sequence to have an allergenic nature, as is the case of *Gly m 8* (**Table 8B**), *Ara h 13.0102*, and *Ara h 15.0101* (**Table 8**: D). Other potential allergens are *Lup a 4* (**Table 8A**), *Lup an 3* and *Lup an 3.0101* (**Table 8A**) and *Lup an delta conglutin*; *Pis s 3, Pis s 3.0101, Pis s 6, Pis s agglutin* and *Pis s albumin* (**Table 8B**); *Ara h 5, Ara h 5.0101* (**Table 8C**), *Ara h 8, Ara h 8.0101, Ara h 8.0102* (**Table 8D**); as both: 43 *Lup l 4* (**Table 8A**); *Ara h 17* (**Table 8D**) and *Cic a 3* (**Table 8C**).

Other proteins assessed as ambiguous or non-allergenic even though they present bibliographic and clinical antecedents of being allergenic include *Lup a gamma* conglutin [34] and *Lup an gamma conglutin* [35] (**Table 8A**); *Ara h 10.0101* [36], *Ara h 11.0101*, and *Ara h 11.0102* [37]; and *Ara i 2.0101* and *Ara i 6.0101* [38] (**Table 8C**).

Allergen name	**T-cell epitopes**	
	LRSSNSFQT	**LRSRNPIYS**
A.		
Range of amino acids occupied by T-cell epitopes joint over soy		
Gly m 5		288–296
Gly m 5.0301	36–44	242–250
Ara h 9.0101	21–29	
Cic a 1		250–258

Allergen name	**T-cell epitopes**					
	LVLVLGIVF	**MMACNGLTI**	**YVLHKIEEI**	**FVLSSSQNS**	**LVAALVLVV**	**LVVHTSASR**
B part 1						
Range of amino acids occupied by T-cell epitopes joint over lupin, peanut, and chickpea						
Lup a 1	11–19					
Lup a 4			66–75			
Lup a alpha conglutin						
Lup a delta conglutin					67–75	73–81
Lup a gamma conglutin		16–24		63–71		
Lup an delta conglutin					62–70	69–77
Lup an gamma conglutin		13–21		77% (FVSSSSQD) 69–77		
Ara d 6	13–20					
Ara h 8.0102			77% (YVLHKIDAI) 66–74			
Cic a 4			88% (YVLHKIEAI) 123–132			

Allergen name	**T-cell epitopes**					
	FQRLNALEP	**LRCAGVALS**	**IRVLERFDQ**	**FGPLRRCN**	**VVLNGRATITI**	**IVRNIKGKN**
B part 2						
Lup a 1			133–138		177–190	
Lup a 4						
Lup a alpha conglutin	83–91	112–120		192–200		279–287
Lup an 1			80% (IRVLERFNQ) 204–212		248–259	

Allergen name	**T-cell epitopes**					
	FQRLNALEP	**LRCAGVALS**	**IRVLERFDQ**	**FGPLRRCN**	**VVLNGRATITI**	**IVRNIKGKN**
Lup an 1.0101			80% (IRVLERFNQ) 204–213		248–260	
Lup an alpha conglutin	86–94	115–123				286–294
Lup an delta conglutin				191–198		
Ara h 1			80% (IRVLQRFDQ) 204–212			
Ara h 1.0101			80% (IRVLQRFDQ) 193–201			

Allergen name	**T-cell epitopes**			
	IVRVSREQI	**IRVNKHM**	**VRRVRRPH**	**WRISDEN**
B part 3				
Lup a 1	302–310			
Lup a alpha conglutin			355–363	
Lup a gamma conglutin		318–326		412–420
Lup an 1	77% (IVRVSKKQI)373–381			
Lup an 1.0101	77% (IVRVSKKQI) 373–381			
Lup an 3.0101			360–367	
Lup an delta conglutin		88% (IRVNKHL) 324–332		88% (WRISSEN) 421–429

Allergen name	**T-cell epitopes**					
	FPILGWLGL	**FVIPAGYPI**	**FVPYYNVNA**	**YVLNGSAWF**	**YVAFKTNDI**	**YKFLVPPPQ**
B part 4						
Lup a 1		433–442				
Lup a 4						
Lup a alpha conglutin	411–418		432–444	445–452	493–501	542–550
Lup an 3.0101	88.88% (FPILRWLGL) 413–421		434–442	447–455	495–503	544–552
Ara h 3			77% (FVPHYNTNA) 404–412			
Ara h 3.0201			77% (FVPHYNTNA) 454–465			

Allergen name	T-cell epitope
	FLLAAHAS
C.	
Range of amino acids occupied by T-cell epitopes joint over peanut	
Ara d 2	13–20
Ara h 2	13–21
Ara h 2.0101	13–21
Ara h 2.0201	13–21
Ara h 2.0202	13–21

This table lists the T -cell epitopes shared on at least two occasions by different species, describing the range of amino acids in which they are located and the percentage of identity with the epitope in the case in which identity is not exact.

Table 7.
Range of amino acids occupied by T-cell epitopes joint over legumes.

Gly m 5, *Gly m 5.0301*, and *Lup a 1* have shown that the IgE epitopes found on these proteins are part of the functional barrel domains of these proteins. In *Lup a 1* protein, two epitopes are located in the Cupin_1.1 domain, which is not affected by post-translational modifications; soybean proteins *Gly m 5* contain an IgE-epitope inside the *Cupin_1* domain, moreover *Gly m 5.0301* also contains the same epitope in the same region and in different positions having no modifications. However, *Gly m 5.0301* does contain epitopes directly affected by glycosidation, within the structural *Cupin_1* domain, an epitope at position 351 (**Table 5A, 6A** and **9A**).

The clinically proven epitopes found in the sequence analysis allowed us to observe how many and to what extent IgE epitopes are shared between proteins of different species and to assess potential cross-reactivity. According to the results, some of the

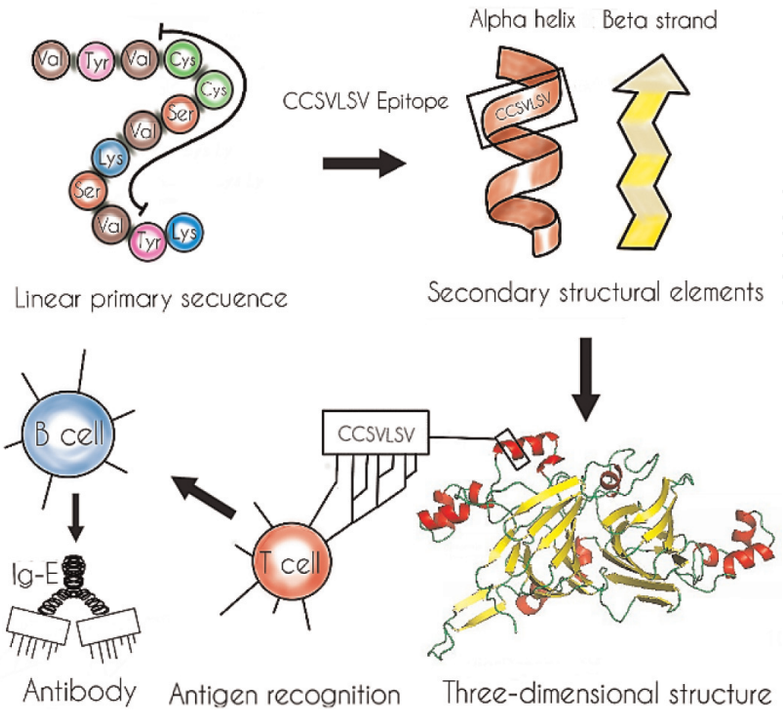

Figure 6.
Summary of the epitope recognision process.

Lupinus angustifolius			***Lupinus albus***		
Allergen name	**Based on amino acid composition**	**Based on dipeptide composition**	**Allergen name**	**Based on amin oacid composition**	**Based on dipeptide composition**
A.					
Prediction of Lupinus allergenic character					
Lup an 1	Potential allergen	Potential allergen	Lup a 1	—	—
Lup an 1.0101	—	—	Lup a 4	Potential allergen	Potential allergen
Lup an 3	Potential allergen	Potential allergen	Lup a alpha conglutin	—	—
Lup an 3.0101	Potential allergen	Potential allergen	Lup a delta conglutin	Potential allergen	Potential allergen
Lup an alpha conglutin	—	—	Lup a gama conglutin	—	—
Lup an delta conglutin	Potential allergen	Potential allergen	***Lupinus luteus***		
Lup an gamma conglutin	—	—	Lup l 4	Allergen	Potential allergen
Pisum sativum			***Glycine max***		
Allergen name	**Based on amino acid composition**	**Based on dipeptide composition**	**Allergen name**	**Based on amino acid composition**	**Based on dipeptide composition**
B.					
Prediction of pea and soy allergenic character					
Pis s 2	Potential allergen	Allergen	Gly m 5	Allergen	Allergen
Pis s 3	Potential allergen	Potential allergen	Gly m 5.0301	Allergen	Allergen
Pis s 3.0101	Potential allergen	Potential allergen	Gly m8	Allergen	Allergen
Pis s 6	Potential allergen	Potential allergen	Gly m 8.0101	Allergen	No allergen
Pis s aglutin	Potential allergen	Potential allergen			
Pis s albumin	Potential allergen	Potential allergen			
Cicer arietinum			***Arachis hypogaea***		
Allergen name	**Based on amino acid composition**	**Based on dipeptide composition**	**Allergen name**	**Based on amino acid composition**	**Based on dipeptide composition**
C.					
Prediction of chickpea and peanut allergenic character					
Cic a 1	—	—	Ara h 1	Allergen	Allergen
Cic a 3	Potential allergen	Allergen	Ara h 1.0101	Allergen	Allergen

Cicer arietinum			*Arachis hypogaea*		
Allergen name	**Based on amino acid composition**	**Based on dipeptide composition**	**Allergen name**	**Based on amino acid composition**	**Based on dipeptide composition**
Cic a 4	Potential allergen	Potential allergen	Ara h 2	—	—
Cic a 6	—	—	Ara h 2.0101	—	—
Arachis duranensis			Ara h 2.0201	—	—
Ara d 2	—	—	Ara h 2.0202	—	—
Ara d 6	—	—	Ara h 3	—	—
Arachis ipaensis			Ara h 3.0201	—	—
Ara i 2.0101	—	—	Ara h 5	Potential allergen	Potential allergen
Ara i 6.0101	—	—	Ara h 5.0101	Potential allergen	Potential allergen
Arachis hypogaea			***A. hypogaea***		
Allergen name	**Based on amino acid composition**	**Based on dipeptide composition**	**Allergen name**	**Based on amino acid composition**	**Based on dipeptide composition**
D.					
Prediction of peanut allergenic character					
Ara h 6	—	—	Ara h 11.0101	—	—
Ara h 6.0101	—	—	Ara h 11.0102	—	—
Ara h 7.0101	Allergen	—	Ara h 13.0102	Allergen	Allergen
Ara h 7.0201	—	—	Ara h 14.0101	—	—
Ara h 7.0301	—	—	Ara h 14.0102	—	—
Ara h 8	Potential allergen	Potential allergen	Ara h 14.0103	—	—
Ara h 8.0101	Potential allergen	Potential allergen	Ara h 15.0101	Allergen	Allergen
Ara h 8.0102	Potential allergen	Potential allergen	Ara h 16	Allergen	—
Ara h 9.0101	Allergen	Allergen	Ara h 17	Potential allergen	Allergen
Ara h 10.0101	—	—	Ara h aglutin	Potential allergen	—

The table summarizes the predictions about the allergenic potential of proteins based on the amino acid and peptide composition. The signal (−) means that the protein has clinically proven epitopes.

Table 8.
Allergenic legume character prediction.

candidate species and proteins for cross-reactivity with soybean (*G. max*) are the peanut (*A. hypogaea*) with three IgE epitopes commonly shared; lupin (*L. albus*) with one epitope in common (**Table 9A**). These findings are supported by bibliographic

Allergen name	**IgE epitopes**					
	HRIFLADKD	**NNFGKLFEVK**	**SYLQEFSRNT**	**ELHLLGFGIN**	**KDLAFPGSGE**	**RRYTARLKEG**
A.						
IgE epitopes shared between different legume species: Glycine max (Gly m), Lupinus albus /Lup a), Arachis hypogaea (Ara h), and Cicer arietinum (Cic a)						
Gly m 5	70% 415-QRNFLAGEKD	70% 297-NNFGKFFEIT	70% 217-SYLQGFSHNI			
Gly m 5.0301	70% 418-QRNFLAGEKD	70% 300-NNFGKFFEIT	70% 220-SYLQGFSHNI			
Lup a 1			70% 286-SYFSGFSRNT	80% 483-NLRLLGFGIN	70% 517-KELTFPGSAE	80% 456-RRYSARLSEG
Lup an 1.0101				80% NLRLLGFGIN	70% KELTFPGSIE	
Ara h 1	100% HRIFLADKD	90% NNFGRLFEVK	90% SYQGFSRNT	100% ELHLLGFGIN	100% KDLAFPGSGE	100% RRYTARLKEG
Ara h 1.0101	100% HRIFLADKD	100% NNFGKLFEVK	100% SYLQEFSRNT	100% ELHLLGFGIN	100% KDLAFPGSGE	100% RRYTARLKEG
Cic a 1				80% DLFLLGFGIN	70% KEVAFPGSAE	

Allergen name	**IgE epitopes**					
	GNIFSGFTPEFLEQA	**IETWNPNNQEFECAG**	**DRRCQSQLER**	**HASARQQWEL**	**KIQRDEDS**	**KRELRNL**
B.						
IgE epitopes shared between different legume species: Lupinus albus (Lup a), Lupinus angustifolius (Lup an), Arachis duranensis (Ara d), Arachis hypogaea (Ara h), and Cicer arietinum (Cic a)						
Lup a alpha conglutin	66.67% GNVLSGFDDEFLEEA	73.34% IETWNPKNDELRCAG				
Lup an alpha conglutin	66.67% GNVLSGFNDEFLEEA	73.34% IETWNPKNDQLRCAG				
Ara d 2			100% DRRCQSQLER	100% HASARQQWEL	100% KIQRDEDS	100% KRELRNL
Ara d 6						85.71% KRELMNL
Ara h 2			100% DRRCQSQLER	100% HASARQQWEL	100% KIQRDEDS	100% KRELRNL
Ara h 2.0101			70% DRRCQSQLER	100% HASARQQWEL	100% KIQRDEDS	100% KRELRNL
Ara h 2.0201			100% DRRCQSQLER	100% HASARQQWEL	100% KIQRDEDS	100% KRELRNL
Ara h 2.0202			100% DRRCQSQLER	100% HASARQQWEL	100% KIQRDEDS	100% KRELRNL
Ara h 3	86.67% GNIFSGFTSEFLAQA	100% IETWNPNNQEFECAG				
Ara h 3.0201	100% GNIFSGFTPEFLEQA	100% IETWNPNNQEFECAG				
Ara h 7.0201						85.71% ERELRNL
Cic a 6	73.33% GNIFSGFKRDFLEDA	73.33% IETWNPSNKQFACAG				

Allergen name	IgE epitopes						
	LQGRQQ	**LRPCEQHLMQ**	**QRCDLDVE**	**QWELQGDR**	**RDPYSP**	**RDPYSP**	**SQDPYSPS**
C.							
IgE epitopes shared between different legume species: Arachis duranensis (Ara d) and Arachis hypogaea (Ara h)							
Ara d 2	100% LQGRQQ	100% LRPCEQHLMQ	100% QRCDLDVE	100% QWELQGDR	100% RDPYSP	83.33% RDPYSP	100% SQDPYSPS
Ara d 6		80% LKPCEQHIMQ	87.5% QRCDLDVS				
Ara h 2	100% LQGRQQ	100% LRPCEQHLMQ	87.5% QRCDLEVE	100% QWELQGDR	100% RDPYSP	83.3% QDPYSP	100% SQDPYSPS
Ara h 2.0101	100% LQGRQQ	100% LRPCEQHLMQ	87.5% QRCDLEVE	100% QWELQGDR	100% RDPYSP	83.3% QDPYSP	100% SQDPYSPS
Ara h 2.0201	100% LQGRQQ	100% LRPCEQHLMQ	87.5% QRCDLEVE	100% QWELQGDR	100% RDPYSP	83.3% QDPYSP	100% SQDPYSPS
Ara h 2.0202	100% LQGRQQ	100% LRPCEQHLMQ	87.5% QRCDLEVE	100% QWELQGDR	100% RDPYSP	83.3% QDPYSP	100% SQDPYSPS
Ara h 6.0101			87.5% QRCDLDVS				
Ara h 7.0201			70% LRPCEEHIRQ				
Ara h 7.0301		70% LRPCEEHIRQ					
D.							
IgE epitopes shared between different legume species: Arachis duranensis (Ara d) and A. hypogaea (Ara h)							
Ara d 2	100% LQGRQQ	100% LRPCEQHLMQ	100% QRCDLDVE	100% QWELQGDR	100% RDPYSP	83.33% RDPYSP	100% SQDPYSPS
Ara d 6		80% LKPCEQHIMQ	87.5% QRCDLDVS				
Ara h 2	100% LQGRQQ	100% LRPCEQHLMQ	87.5% QRCDLEVE	100% QWELQGDR	100% RDPYSP	83.3% QDPYSP	100% SQDPYSPS
Ara h 2.0101	100% LQGRQQ	100% LRPCEQHLMQ	87.5% QRCDLEVE	100% QWELQGDR	100% RDPYSP	83.3% QDPYSP	100% SQDPYSPS
Ara h 2.0201	100% LQGRQQ	100% LRPCEQHLMQ	87.5% QRCDLEVE	100% QWELQGDR	100% RDPYSP	83.3% QDPYSP	100% SQDPYSPS
Ara h 2.0202	100% LQGRQQ	100% LRPCEQHLMQ	87.5% QRCDLEVE	100% QWELQGDR	100% RDPYSP	83.3% QDPYSP	100% SQDPYSPS
Ara h 6.0101			87.5% QRCDLDVS				
Ara h 7.0201		70% LRPCEEHIRQ					
Ara h 7.0301		70% LRPCEEHIRQ					

This table summarizes the IgE epitopes clinically confirmed in different species, and the accuracy percentage of these epitopes found according to the protein sequence.

Table 9.
IgE epitopes shared between different legume species.

reports [38]. It is also found that *L. albus* shares four epitopes with *A. hypogea* and *L. angustifolius* (**Table 9A**), and other two with *A. hypogea*. Looking at other cases, it is observed that in close species such as peanut, species such as *A. duranensis* and *A.*

Allergen name	IgE epitopes			
	DITNPINLRE	**KESHFVSARP**	**EQEERGQRRW**	**VTVRGGLRILSPDRK**
IgE epitopes shared only by same legume species: Arachis hypogaea (Ara h)				
Ara h 1	90% DITNPINLRD	90% RESHFVSARP	90% EQEERGQRR	
Ara h 1.0101	100% DITNPINLRE	100% KESHFVSARP	100% EQEERGQRRW	
Ara h 3				93,345% VTCRGGLRILSPDRK
Ara h 3.0201				86.67% VTVRGGLRILSPDRK

This table summarizes the IgE epitopes clinically confirmed in different species, and the accuracy percentage of these epitopes found according to the protein sequence.

Table 10.
IgE epitopes shared only by same legume species.

hypogea shared ten common epitopes (**Table 9B, C**), similarly to *Lupinus* finding four epitopes in common (**Table 9A, B**).

In addition, shared T-cell epitopes have been found among species that do not include soybean such as *L. albus* and *L. angustifolia* (**Table 9**: AB), but not found in *L. luteus*; *A. hypogaea* (**Table 9A-D**), and *A. duranensis* (**Table 9B-D**); *C. arietinum* (**Table 9A, B**); and *P. sativum* (**Table 9A**). These epitopes have been identified as relevant epitopes in previous studies on sensitizations between allergens of different species with similar structure and sequence leading to the development of allergic cross-reactions [38, 39].

An interesting fact is that different isoforms of the same protein may or may not present the same IgE epitope and, in the case of having it, it does not necessarily have the same degree of similarity. Establishing a relationship with the information obtained in the alignments, we can conclude that the small differences observed in the sequence between isoforms of the same protein can be key to conformation and epitopes presence (**Table 10**).

4. Conclusions

This chapter presented a study of functional and allergenic features of legume seed proteins.

Analysis of allergenic legume proteins legume as well as all available isoforms allowed for extracting shared epitopes that can be linked to cross-reactivity processes among the eight studied species (*G. max, A. hypogaea, L. albus, L. angustifolius, A. duranensis, C. arietinum, P. sativum,* and *L. culinaris)*. Shared epitopes were not found with soybean or with the rest of the legume allergens examined from *A. duranensis*.

Small differences in the amino acid sequences (less than 1%) of the same allergen isoforms implied important changes in epitopic conformation and sequences of T-cell and IgE recognizable epitopes. Small differences in amino sequences of isoforms from the same inferred changes over 2D and 3D structure conformation that may affect

functional protein domains. Post-translational modifications allowed identification of possible phosphorylation, glycosylation, carboxylation, methylation, nitrosylation, and nitration sites in protein functional domains, near or directly located in different type of epitopes with potential influence in allergenic response.

Primary sequence alignments together with three-dimensional protein modeling allowed to study the conservation of proteins as conglutin gamma proteins among different *Lupinus*. species, assessing also their potential allergenicity.

The changes described close to the sequence or related to spatial distribution of the epitopes may involve potential alterations on protein allergenicity.

Obtaining reliable clinical data on legume allergies in developing countries could be helpful in clarifying whether the increase in food allergies is actually due to poor dietary habits and increasing industrialization processes.

Further studies on the characterization of more allergenic proteins, including isoforms of major allergens already described, not only sequential but also three-dimensional conformational epitopes, can be a great advancement for the prevention of cross-reactivity and the improvement of knowledge of allergies produced by legumes, which in turn could promote the introduction of this food as a substitute for other foods of lower nutritional quality and with greater environmental impact.

Acknowledgements

This study has been partially funded by The Spanish Ministry of Economy, Industry and Competitiveness through the grants Ref.: RYC-2014-2016,536 (Ramon y Cajal Research Program) to JCJ-L; and Ministry of Health and Families, Andalusian government. Funding for I + D + i in biomedical research and health sciences in Andalusia, grant Ref.: PI-0450-2019.

Conflicts of interest

The authors have declared that no competing interests exist.

Abbreviations

LTP	Lipid Transfer Protein
3D	three-dimensional
PR	proteins Pathogenesis-related proteins

DOI: http://dx.doi.org/10.5772/intechopen.106971

Author details

Andrea Roman-Mateo[1], Esther Rodriguez-de Haro[1], Jose M. Berral-Hens[1], Sonia Morales-Santana[2] and Jose C. Jimenez-Lopez[1,3]*

1 Department of Biochemistry, Cell and Molecular Biology of Plants, Estación Experimental del Zaidín, Spanish National Research Council (CSIC), Granada, Spain

2 Proteomic Research Department, San Cecilio University Hospital, Biosanitary Research Institute of Granada (Ibs.GRANADA), Granada, Spain

3 The UWA Institute of Agriculture and School of Agriculture and Environment, The University of Western Australia Perth, WA Australia

*Address all correspondence to: josecarlos.jimenez@eez.csic.es

References

[1] Myles IA. Fast food fever: Reviewing the impacts of the Western diet on immunity. Nutrition Journal. 2014;**13**. Article number 61:2-4

[2] Tang MLK, Mullins RJ. Food allergy: Is prevalence increasing? Internal Medicine Journal. 2017;**47**:256-261

[3] Gray CL, Levin ME, du Toit G. Ethnic differences in peanut allergy patterns in south African children with atopic dermatitis. Pediatric Allergy and Immunology. 2015;**26**: 721-730

[4] Rani K, Sharma P, Kumar S,Wati L, Kumar R, Gurjar DS, et al. Legumes for Sustainable Soil and Crop Management. In: Meena R, Kumar S, Bohra J, Jat M. (eds) Sustainable Management of Soil and Environment. Singapore: Springer; DOI: 10.1007/978-981-13-8832-3_6

[5] Stagnari F, Maggio A, Galieni A, et al. Multiple benefits of legumes for agriculture sustainability: An overview. Chemical and Biological Technologies in Agriculture. 2017;**4**:2

[6] Hildebrand HV, Arias A, Simons E, Povolo B, Janet Rothney MLIS, Protudjer JLP. Adult and pediatric food allergy to chickpea, pea, lentil, and lupine: A scoping review. The Journal of Allergy and Clinical Inmmunology. 2020;**9**(1):290-301

[7] Scott H, Sicherer MD. Clinical implications of cross-reactive food allergens. Journal of Allergy and Clinical Immunology. 2001;**108**(6):881-890

[8] Jappe U, Vieths S. Lupine, a source of new as well as hidden food allergens. Molecular Nutrition Food Research. 2010;**54**:113-126

[9] San Ireneo MM, Ibáñez MD, Sánchez J-J, Carnés J, Fernández-Caldas E. Clinical features of legume allergy in children from a Mediterranean area. Annals of Allergy, Asthma and Inmunology. 2008;**101**(2):179-184

[10] Alvarado MI, Pérez M. Study of food allergy on Spanish population. Allergologia et Immunopathologia. 2006;**34**(5):185-193

[11] Richard C, Jacquenet S, Sergeant P, Moneret-Vautrin DA. Cross-reactivity of a new food ingredient, dun pea, with legumes, and risk of anaphylaxis in legume allergic children. European Annals of Allergy and Clinical Immunology. 2015;**474**(4):118-125

[12] Nwaru BI, Hickstein L, Panesar SS, Muraro A, Werfel T, Cardona V. On behalf of the EAACI Food Allergy and Anaphylaxis Guidelines Group. The epidemiology of food allergy in Europe: A systematic review and meta-analysis. Allergy. 2014;**69**:62-75

[13] Tham EH, Leung DYM. How different parts of the world provide new insights into food allergy. Allergy Asthma Immunol research. 2018;**10**(4):290-299

[14] Gaiaschi R, Beretta P, Fiocchi C, Velonà P, Galli U. Cross-reactivity between milk proteins from different animal species. Clinical & Experimental Allergy. 2001;**9**(7):997-1004

[15] Lallès JP, Peltre G. Lead review article biochemical features of grain legume allergens in humans and animals. Nutrition Reviews. 1996;**54**(4):101-107

[16] Mills ENC, Jenkins J, Marigheto N, Belton PS, Gunning AP, Morris VJ. Allergens of the Cupin superfamily.

Biochem Society Transactions. 2002; **30**(6):925-929

[17] Asero R, Mistrello G, Roncarolo D, de Vries S, Gautier M, Ciurana CL, et al. Lipid transfer protein: A pan-allergen in plant-derived foods that is highly resistant to pepsin digestion. International Archive of Allergy Immunology. 2001;**124**:67-69

[18] Wang D, Liu D, Yuchi J, He F, Jiang Y, Cai S, et al. MusiteDeep: A deep-learning based webserver for protein post-translational modification site prediction and visualization. Nucleic Acids Research. 2020;**48**(1):140-146

[19] Kattan JD, Sampson HA. Clinical reactivity to soy is best identified by component testing to Gly m 8. Journal Allergy and Clinical Immunoly. 2015; **3**(6):970-972

[20] Ueberham E, Spiegel H, Havenith H, Rautenberger P, Lidzba N, Schillberg S, et al. Simplified tracking of a soy allergen in processed food using a monoclonal antibody-based Sandwich ELISA targeting the soybean 2S albumin Gly m 8. Journal of Agricultural and Food Chemistry. 2019;**67**(31):8660-8667

[21] Maruyama N, Sato S, Cabanos C, Tanaka A, Ito K, Ebisawa M. Gly m 5/Gly m 8 fusion component as a potential novel candidate molecule for diagnosing soya bean allergy in Japanese children. Clinical and Experimental Allergy. 2018; **48**:1726-1734

[22] CODEX Alimentarius Commission. Proposed draft annex on the assessment of possible allergenicity of the draft guideline for the conduct of food safety assessment of foods derived from recombinant DNA plants. Joint FAO/ WHO Food Standard Program. Appendix IV. 2003:57-60

[23] Midun E, Radulovic S, Brough H, Caubet J-C. Recent advances in the management of nut allergy. World Allergy Organization Journal. 2021; **14**(1):1939-4551

[24] Bublin M, Breiteneder H. Cross-reactivity of Peanut allergens. Current Allergy and Asthma Reports. 2014;**14**: 426

[25] McGinty JW, Marré ML, Bajzik V, et al. T cell epitopes and post-translationally modified epitopes in type 1 diabetes. Currently Diabetes Repport. 2015;**15**:90

[26] Narayanan A, Jacobson MP. Computational studies of protein regulation by post-translational phosphorylation. Current Opinion in Structural Biology. 2009;**19**(2):156-163

[27] Deribe Y, Pawson T, Dikic I. Post-translational modifications in signal integration. Nature Structural & Molecular Biology. 2010;**17**:666-672

[28] Mueller GA, Maleki SJ, Johnson K, Hurlburt BK, Cheng H, Ruan S, et al. Identification of Maillard reaction products on peanut allergens that influence binding to the receptor for advanced glycation end products. Allergy. 2013;**68**:1546-1554

[29] Kurotani A, Tokmakov AA, Kuroda Y, Fukami Y, Shinozaki K, Sakurai T. Correlations between predicted protein disorder and post-translational modifications in plants. Bioinformatics. 2014;**30**(8): 1095-1103

[30] Astier J, Rasul S, Koen E, Manzoor H, Besson-Bard A, Lamotte O, et al. S-nitrosylation: An emerging post-translational protein modification in plants. Plant Science. 2011;**181**(5): 527-533

[31] Corpas FJ, Begara-Morales JC, Sánchez-Calvo B, Chaki M, Barroso JB. Nitration and S-Nitrosylation: Two Post-translational Modifications (PTMs) Mediated by Reactive Nitrogen Species (RNS) and Their Role in Signalling Processes of Plant Cells. In: Gupta K, Igamberdiev A. (eds) Reactive Oxygen and Nitrogen Species Signaling and Communication in Plants. Signaling and Communication in Plants, vol 23. Cham: Springer. 2015. DOI: 10.1007/978-3-319-10079-1_13

[32] Dunwell JM, Culham A, Carter CE, Sosa-Aguirre CR, Goodenough PW. Evolution of functional diversity in the Cupin superfamily. Trends in Biochemical Sciences. 2001;**26**(12): 740-746

[33] Jimenez-Lopez JC, Rodríguez-García MI, Alché JD. Analysis of the Effects of Polymorphism on Pollen Profilin Structural Functionality and the Generation of Conformational, T- and B-Cell Epitopes. PLOS ONE. 2013;**8**(10): e76066

[34] Svobodova M, Mairal T, Nadal P, Bermudo MC, O'Sullivan CK. Ultrasensitive aptamer based detection of β-conglutin food allergen. Food Chemistry. 2014;**165**:419-423

[35] Nadal P, Svobodova M, Mairal T, O'Sullivan CK. Probing high-affinity 11-mer DNA aptamer against Lup an 1 (β-conglutin). Analytical and Bioanalytical Chemistry. 2013;**405**(29):9343-9349

[36] Cabanos C, Katayama H, Tanaka A, Utsumi S, Maruyama N. Expression and purification of peanut oleosins in insect cells. Protein Journal. 2011;**30**(7): 457-463

[37] Ramos ML, Fleming G, Chu Y, Akiyama Y, Gallo M, Ozias-Akins P. Chromosomal and phylogenetic context for conglutin genes in Arachis based on genomic sequence. Molecular Genetics and Genomics. 2006;**275**(6):578-592

[38] Cabanillas B, Jappe U, Novak N. Allergy to Peanut, Soybean, and Other Legumes: Recent Advances in Allergen Characterization, Stability to Processing and IgE Cross-Reactivity. Molecular Nutrition and Food Research. 2018;**62**(1)

[39] Verma AK, Kumar S, Das M, et al. A comprehensive review of legume allergy. Clinical Reviews in Allergy & Immunology. 2013;**4**:30-46