Structure and IgE cross-reactivity among walnut and peanut vicilin leader sequences

Jacqueline Nesbit[1], Alexander Foo[2], Stephen Gipson[1], Pierre Bushel[3], Eugene DeRose[3], Hsiaopo Cheng[1], Catherine Schein[4], Suzanne Teuber[5], Barry Hurlburt[1], Geoffrey Mueller[3], and Soheila Maleki[1]

[1]USDA
[2]National Institute of Environmental Health Sciences
[3]NIEHS
[4]UTMB
[5]UC Davis

June 23, 2021

Abstract

Background: Vicilin seed storage proteins are translated with N-terminal leader sequences (LSs) that are cleaved to yield the mature protein. These LSs were thought to be unstructured and rapidly degraded. However, Ara h 1 and Jug r 2 LS (A1LS, J2LS) have been identified in seeds, and immunodominant IgE epitopes detected. Here, common sequences containing structured CxxxC-repeat motifs were identified as potential mediators of IgE cross-reactivity despite very low (17%) sequence identity. Method: Linear IgE epitopes were identified by peptide microarrays, in which overlapping 15-mer peptides on glass slides, were incubated with sera from peanut, walnut or dual allergic individuals. Similar epitopes were computationally predicted. Peanut A1LS and walnut J2LS fragments (J2.1, J2.2, J2.3) each with a CxxxC vicilin LS motif were identified, cloned, expressed, purified and their structures solved using solution-NMR to locate and assess epitopes on the structure. Results: A1LS and J2LSs reveal similar helix-turn-helix motifs connected by disulfide bonds between adjacent CxxxC repeats forming ∝-hairpin structures. Peanut-allergic IgE bound more frequently to the J2LSs, regardless of walnut allergic status or A1LS binding. IgE binding pattern to peptides from both J2LS and A1LS, along with structure and computational predictions, suggest that the structure and conserved amino acid properties of peptides determine cross-reactivity. The properties of LS IgE epitopes were closely related to epitopes in 2S albumins. Conclusion: The shared ∝-hairpin structure is a stable scaffold that contributes to cross-reactivity despite low sequence identity. Biophysical properties are a better predictor of distant cross-reactivity than traditional measures of evolutionary conservation.

Structure and IgE cross-reactivity among walnut and peanut vicilin leader squences

Running title: Vicilin leader sequences contribute to cross-reactivity

Jacqueline B. Nesbit[*1], Alexander C.Y. Foo[*2], Stephen A.Y. Gipson[1], Pierre Bushel[2], Eugene F. DeRose[2], Hsiaopo Cheng[1], Catherine H. Schein[3], Suzanne S. Teuber[4], Barry K. Hurlburt[1], Geoffrey A. Mueller[+2], Soheila J. Maleki[+1]

1- USDA-ARS

2- National Institute of Environmental Health Sciences

3- Department of Biochemistry and Molecular Biology, Institute for Human Infections and Immunity, University of Texas Medical Branch, 301 University Blvd., Galveston TX 77555

4- University of California Davis School of Medicine

* Denotes equal contributions

+ Corresponding authors:

Soheila.Maleki@usda.gov

1100 Robert E Lee Blvd.

New Orleans, LA 70124

504-286-4590

Geoffrey.Mueller@nih.gov

111 T.W. Alexander Dr. MD-MR01

Research Triangle Park, NC 27709

Authors contributions : JBN Performed the microarray analysis and wrote a significant part of the manuscript; ACYF performed a majority of the structural experiments and wrote a significant part of the manuscript; EFD and GAM assisted with structural studies. SAYG performed the statistical analysis with guidance from PB. BKH, HC designed and performed the microarray experiments; CHS performed the PD computational work; SJM and GAM co-directed the study design and implementation and, helped prepare figures and manuscript.

Acknowledgements:

SAXS data was collected at SIBYLS which is supported by the DOE-BER IDAT DE-AC02-05CH11231 and NIGMS ALS-ENABLE (P30 GM124169 and S10OD018483). This research was supported in part by the Intramural Research Program of the National Institute of Environmental Health Sciences, (Z01-ES102906, GAM) and the National Institute of Allergy and Infectious Disease NIAID-R21 AI135397. Also funded by the US Department of Agriculture: Agricultural Research Service (ARS-6054-43440-044-00D) and the National Institute for Food and Agriculture/National Peanut Board (NIFA-60-6054-7-003 SJM/BKH/JBN). The structures for A1LS, J2.1, J2.2, and J2.3 submitted to the Protein Data Bank and the Biological Magnetic Resonance Bank under the access codes PBD ID 7LXK BMRB 30875, PDB ID 7LVF BRMB ID 30870, PDB ID 7LVG BMRB 30871, and PDB ID 7LVE BRMB ID 30869 respectively.

Abstract:

Background: Vicilin seed storage proteins are translated with N-terminal leader sequences (LSs) that are cleaved to yield the mature protein. These LSs were thought to be unstructured and rapidly degraded. However, Ara h 1 and Jug r 2 LS (A1LS, J2LS) have been identified in seeds, and immunodominant IgE epitopes detected. Here, common sequences containing structured CxxxC-repeat motifs were identified as potential mediators of IgE cross-reactivity despite very low (17%) sequence identity.

Method: Linear IgE epitopes were identified by peptide microarrays, in which overlapping 15-mer peptides on glass slides, were incubated with sera from peanut, walnut or dual allergic individuals. Similar epitopes were computationally predicted. Peanut A1LS and walnut J2LS fragments (J2.1, J2.2, J2.3) each with a CxxxC vicilin LS motif were identified, cloned, expressed, purified and their structures solved using solution-NMR to locate and assess epitopes on the structure.

Results: A1LS and J2LSs reveal similar helix-turn-helix motifs connected by disulfide bonds between adjacent CxxxC repeats forming α-hairpin structures. Peanut-allergic IgE bound more frequently to the J2LSs, regardless of walnut allergic status or A1LS binding. IgE binding pattern to peptides from both J2LS and A1LS, along with structure and computational predictions, suggest that the structure and conserved

amino acid properties of peptides determine cross-reactivity. The properties of LS IgE epitopes were closely related to epitopes in 2S albumins.

Conclusion: The shared α-hairpin structure is a stable scaffold that contributes to cross-reactivity despite low sequence identity. Biophysical properties are a better predictor of distant cross-reactivity than traditional measures of evolutionary conservation.

Κεψ ωορδς: α-ηαιρπιν σςαφφολδ, ςροσσ-ρεαςτιιτψ, ςψστεινε ρεπεατ μοτιφ, IγE επιτοπες, λεαδερ σεχυενςε

Background:

Peanut (PN) allergy is among the most commonly reported food allergies, with a global prevalence of 1-3% [1]; a number that is only increasing. Peanut allergy is responsible for over 50% of food-induced anaphylaxis cases, and is rarely outgrown, making it a key health concern [2, 3]. Furthermore, up to 86% of peanut-allergic patients show some level of sensitization to tree-nuts (TN), particularly walnut (WN), with ˜30% displaying clinically-relevant levels of IgE cross-reactivity [4, 5]. While allergens from PNs and TNs can be grouped into similar protein families, their overall sequence identity falls well below the 70% threshold previously considered to support cross-reactivity, raising questions regarding the basis for this phenomenon [6]. Dreskin et al. [7] suggest that the highly conserved cysteine motif structure of 2S albumins, which supports intervening amino acids with similar physicochemical properties (PCPs), is essential for IgE cross-reactivity between PN and TNs. The cysteine-rich leader sequences (LS's) from viclin seed storage proteins could potentially harbour similar interactions, further contruibuting to PN/TN cross-reactivity. Indeed, peptides from the vicilin leader sequences (LS) of walnut and peanut and that have similar PCPs to the 2S-albumin allergens of peanut (Ara h 2, Ara h 6 and Ara h 7) show comparable IgE binding properties [8, 9], supporting this conjecture.

Vicilins are a family of ubiquitous seed storage proteins which are translated as preproproteins consisting of three conserved regions: a short, hydrophobic signal peptide (pre) that is removed upon location to a storage vacuole; an N-terminal, cysteine-rich pro region; and the mature vicilin domain [10-14]. The pro region is processed by an asparaginyl endopeptidase (AEP) [15] to yield the leader sequence (LS). The LS in turn is composed of a contain a variable number of vicilin-buried peptide (VBP) sequences, defined by a common cysteine motif (CxxxC(10-14)CxxxC). The native LSs from both WN vicilin, Jug r 2(J2LS) and PN vicilin, Ara h 1 (A1LS) have been identified in the extracts of mature nuts [16, 17] and shown to bind IgE from patients with allergy to PN or TN [9, 17]. Despite their taxonomical and sequence diversity, VBP's adopt a common α-hairpin structure mediated by disulfide bonds between the highly conserved CxxxC motifs. While none of the VBP's described in these studies are known to be immunogenic, this sequence-insensitive scaffold has the potential to elicit an immune response. It is possible that this scaffold contains cross-reactive epitopes with similar PCPs across distantly related, IgE-reactive proteins from different plat species, such as PN and WN.

The contribution of PCPs to IgE cross-reactivity was analyzed by Maleki, et al. [8] in which peptides in the J2LS, with similar PCPs to an immunodominant epitope of the peanut 2S albumin, Ara h 2 (DRRCQSQLER), were shown to correlate with clinically relevant cross-reactivity despite low sequence identity. Indeed, the epitopes identified in the walnut LS bound IgE from PN and TN allergic individuals more intensely than the original immunodominant epitope of Ara h 2. In addition, a consensus peptide (CQRQEQGQRQQQQ) was designed based on the alignment of Ara h 2, Ara h 6 and Ara h 7 with the repeats of CxxxC motif within the J2LS. Using the peptide similarity tool in the Structural Database of Allergenic Proteins (SDAP) [18-21], Nesbit et al. identified IgE binding repeats with PCPs similar to the consensus in a multitude of other nut allergens' sequences [9]. Anti-consensus peptide antiody was shown to recognize vicilins, glycinins and 2S albumins in different nuts. The J2LS and a 31 mer containing the consensus sequence were both shown to inhibit IgE binding to walnuts and almond extracts. While these efforts establish J2LSs from walnut as potential mediators of cross-reactivity with 2S albumins, its cross-reactivity with LS's

from PN and other distantly related food sources have yet to be explored in a comprehensive manner.

Here, we show by solution NMR that the A1LS and the J2LS adopt a common α-hairpin fold conferred by disulfide bonding between the cysteine residues of the CxxxC motifs. Despite their ordered structure, few inter-helical interactions were observed beyond the conserved disulfide bonds, allowing LSs to tolerate significant sequence variation with minimal structural perturbation. Peptide microarray analysis containing the A1LS and the J2LS reveals the potential for a high degree of IgE cross-reactivity among these LSs [9]. Though these specific immunogenic properties cannot be easily assessed using traditional bioinformatic approaches, a combination of structural, computational and immunological approaches provides insights into the joint contribution of structure and sequence to clinical cross-reactivity.

Methods:

Methods related to the constructs, purification, biophysical characterisation, NMR, and microarray statistical analysis are located in the supplemental information.

Patient sera

Sera from 40 individuals with walnut (12 patients), peanut (12 patients), and peanut/walnut allergy (16 patients) were collected in accordance with rules and regulations of the institutional review board of their respective institutions (Tulane University Biomedical IRB 09-00231, REF #: 140613; University of California at Davis IRB protocol No. 200210194-6) and in accordance with U.S. federal policy for the protection of human subjects (Table 1). Patients were over the age of 18 and had experienced recurrent severe, systemic allergic reactions to peanut, walnut or both. These patients were included in this study based on their convincing clinical history of food allergy and did not undergo food challenge due to the potential severity of reactions.

Microarray

The entire amino acid sequences of Ara h 1 and Jug r 2 were printed onto microarray slides (JPT Peptide Technologies GmbH, Berlin, Germany) as sequential overlapping 15 amino acid spots, offset by 5 amino acids. Slides were placed into a HS400 Pro (Tecan, San Jose, CA), blocked in filtered SuperBlock TBS (Thermo Fisher Scientific, Waltham, MA) for 30 minutes at room temperature under agitation and then washed with Tris-buffered saline containing 0.5% Tween-20 (TBST) (Bio-Rad, Hercules, California). After centrifugation, 200μL of each patient's undiluted sera was injected into individual chambers containing microarray slides and incubated at 4°C for 16 hours with agitation. Microarray slides were then washed and injected with 170 μL of mouse α-human IgE (Life Technologies, Grand Island, NY) diluted into filtered Superblock at a dilution of 1:5000 for 30 minutes at room temperature. The slides were washed and dried before scanning on a GenePix-4000B (Molecular Devices, San Jose, CA). IgE binding was measured by Cy3 green fluorescence at 532nm. The data was analyzed by GenePix Pro 7.2 software.

Statistics

Statistical analyses were performed using R (version 3.6.3). Modified z values were calculated from the microarray median signal-to-noise ratios (SNRs) and used to determine the percent of leader sequence peptides bound by IgE for each allergy group. Median SNR values were converted into modified z values, calculated using median and median absolute deviation (MAD) of a patient's SNRs to an entire leader sequence or leader sequence fragment. To approximate standard deviation, MAD was calculated using the constant of 1.4826. We defined a positive binding event as a z value $>= 3$. Using the *fisher.test*function, Fisher's exact tests were used to compare the proportion of positive binding events out of the total possible binding events for a particular fragment among allergy groups. When Fisher's tests indicated a significant difference between one or more allergy groups (p-value < 0.05), pairwise post hoc comparisons were made to determine

which allergy groups differed from one another. We defined major epitopes as peptides positively bound by 50% or more or patients within an allergy group. Average z values which directly relate to IgE binding intensity are presented as low (3 – 6), medium (6.1 – 9) and high (>9). See supplemental methods for more details.

NMR

Detailed NMR methods can be found in supplemental material. Briefly, triple-resonance and NOESY spectra were collected on 0.1-1 mM protein samples in PBS using either a 600 or 800 MHz Agilent DD2 console equipped with cryogenically cooled probe. Backbone and side-chain assignments and T1/T2 relaxation times for the oxidized samples were obtained using standard triple resonance techniques employing either the standard VARIAN Biopack or modified BEST-TROSY pulse sequences as described previously[22-25].

Property distance graph (PD-graph)

The PD-graph calculates the interpeptide PD for a group of sequences and graphically presents their similarity to one another at the PCP level [26]. The program produces a graphic optimizing a metric of how the distances between peptides relate to one another using gradient descent from random initial placement of the sequences. First it automatically calculates the pairwise interpeptide physicochemical property (PCP) similarity (PD, where identical sequences have a PD of 0 and values increase with increasing divergence) of a given set of sequences and then determines through an iterative process the best clustering. PD-graph produces reproducible graphics using a helper script that normalizes the orientation (rotation and reflection) of each graphic to show a consensus of multiple runs. The graph labeling and other features are customized by the user.

Results:

ὁμμονα-ηαιρπιν φολδ προιδες α στρυςτυραλ βασις φορ IγE ςροσσ-ρεαςτιιτψ

Examining the A1LS sequence reveals a single VBP motif with an additional pair of cysteines at both N and C-terminals, which likely fold to form a single unit. In contrast, the JR2LS contains three VBP motifs (Figure 1A and 1B). The J2LS degrades into a series of 5-7 kDa fragments corresponding to the individual α-hairpins [16]. Based on this background, the three J2LS VBP motifs (J2.1, J2.2, J2.3) were produced recombinantly and their structures analyzed individually. Circular dichroism (CD) spectra of all four VBPs (A1LS, J2.1, J2.2, J2.3) show distinct minima at 222 and 210 nm consistent with predominantly α-helical secondary structure (S1). Further studies on the full-length J2LS using small-angle X-ray scattering (SAXS) yielded a radius of gyration of 25±2Å; significantly greater than the 20Å expected for a globular protein of equivalent molecular weight. Additionally, the shape of the corresponding Kratky plot suggested a high degree of flexibility. This data supports the proposed model in which J2.1, J2.2, and J2.3 could exist as independently folded hairpins separated by unstructured linker regions (S2), and validates our approach in examining each of these LSs as separate entities rather than part of a larger globular system.

The structural similarity afforded by the common α-hairpin scaffold could potentially give rise to cross-reactive epitopes despite the lack of sequence identity (Figure 1C). To explore this possibility, the structure of the WN and PN LSs was assessed using solution-NMR. Backbone and side-chain assignments were obtained using standard triple-resonance approaches. Analysis of the resulting chemical shifts using the TALOS prediction algorithm reveals two α-helical regions centered on the CxxxC motifs (S2). The downfield shift (>33 ppm) of the cysteine C_b peaks, along with the presence of NOESY cross-peaks between the C_b protons on opposite CxxxC repeats support the presence of an α-hairpin disulfide bonding pattern (S3). The A1LS contains an additional pair of cysteine residues located on either side of the main CxxxC motif (Figure 1B), which were found to form a third disulfide bond that was verified by mass-spectrometry (S1), explaining the high stability and survival of the intact form in the seed. This information, along with structural restraints derived from the available NMR data was used to determine the 3D structure of all four LS fragments. As shown in Figure 1A, all four constructs adopt an unambiguous α-hairpin structure with low backbone RMSD values (S3). Few long-range NOE interactions were detected beyond the disulfide bonded cysteines

in each of the 4 LS fragments; a rather unusual observation for an ordered protein (S3). To further verify the calculated structures, rotational correlation times (T_c) were measured for all constructs (Figure 1A). The resulting values are consistent with a >7 kDa globular protein and not with an unfolded peptide. Taken together, this data suggests that A1LS, J2.1, J2.2 and J2.3 each adopt a common α-hairpin structure maintained by the conserved disulfide bonds. This architecture is common to the VBPs from other plant species [14, 27, 28], and provides a consistent scaffold that allows for permutations in amino acid sequence as shown by the low sequence identity beyond the conserved CxxxC motifs.

Independently of the IgE binding analysis, the ability of these α-hairpin structures to support cross-reactive epitopes was evaluated using the SPADE computational tool developed by Dall'Antonia *et al.* [29]. The A1LS amino acid sequence is shown (Figure 2A and 2B) with previously identified epitopes and the location of the overlapping peptides on the structure, respectively. The SPADE algorithm takes into account the physical properties of each residue similar to the PD metric employed for sequence analysis, but includes an additional layer using the 3D structure to examine solvent accessible area and spatial positioning from which regions of similarity corresponding to putative cross-reactive 3D epitopes can be identified. Comparing the structure of A1LS with the three J2LS VBP's reveal areas of high surface similarity (ΔSIM) (Figure 2C), representing regions of potential cross-reactivity. However, this similarity was markedly reduced in J2.3 despite the conserved α-hairpin fold. The reverse comparison yields similar results (Figure 3A), with JR2.3 once again displaying noticeably lower overall (ΔSIM) values against A1LS than its stablemates. The amino acid sequence of the J2LS VBPs and the location of the overlapping peptides are shown in Figure 3B and 3C.

IgE binding to A1LS and J2LS

Peptide microarrays were used to measure IgE binding to J2LS and A1LS peptides using sera of patients with allergy to PN, WN or both PW. First, we compared IgE binding (Figure 4A) to the overlapping peptides of the long isolated VBP repeats characterized by NMR (Figure 1). The majority of IgE binding in our patient cohort was to the peptides of A1LS and J2.1 (Figure 4A), consistent with surface similarity values shown in Figure 3A. IgE from the sera of patients with PW and PN allergic only bind equally well to J2.1 and A1LS, while IgE from WN only sera was statistically significantly lower (i.e., PN = PW > WN). However, IgE from WN allergic also bound the A1LS, further supporting the role of leader sequences in allergic cross-reactivity.

Next, IgE binding to overlapping 15 amino acid peptides from the LSs was analyzed to identify epitopes recognized at least one allergy group (WN, PN and PW). Nine immunoactive peptides were identified, three in A1LS (A4, A12 and A13), five in J2.1 (JR5, JR6, JR7, JR8, and JR9), and one in J2.2 (JR18) (Figures 2A, 3B and 4B). The binding pattern suggested that in some cases the peptides represent fragments of longer, and potentially conformational epitopes. All the IgE binding peptides identified in A1LS contain previously identified immunodominant epitopes of Ara h 1 [30, 31]. A4, located in a disordered region of A1LS (Figure 2C), was only recognized by PN allergic sera. A12 spans the most structured portion of A1LS and binds IgE from both PN and PW sera, while A13, contains both ordered and disordered segments, is recognized by all three serum groups (PN, WN, PW). Both A12 and A13 overlap with the immunodominant epitope, LEYDPRCVYD (Figure 2A)[32] though this sequence is truncated in A12, resulting in a reduction in both the number of sera recognizing it and the intensity of IgE binding based on average z values (Figure 4B). Although peptides A10 and A11 both contain the A1LS epitope, QEPDDLKQKAC, the latter has no IgE binding, suggesting that the surrounding sequences and possibly conformation in reactivity.

As Figure 4 shows, JR6, JR7, JR8 and JR9, bind IgE in sera from all three patient groups, and all are peptides of JR2.1 that would remain intact after trypsin digestion (Fig. 1B). This suggests that they could contribute to cross-reactions between walnut and peanut. We compared the physicochemical property (PCP) similarity of these peptides to other known IgE epitopes using a PD-graph, which groups peptides according to their similarity in PCPs (Figure 5) [26]. Low PD values between peptides that have been predicted to be similar to known IgE epitopes have previously been shown to have IgE reactivity [8, 9]. Figure 5 illustrates the most similar peptide sequences as connected by blue lines, and the lines descrease in intensity and width

with increasing PD value (lower similarity). These peptides are similar to each other and to the IgE epitopes of the 2S albumins and the consensus peptides defined previously [8, 9]. Peptides JR6, 7, 8, 9 and A13 (shown in red boxes) have some of the highest z scores and are clustered in the PD graph, which is consistent with the SPADE similarity parameter results for JR2.1 and A1LS.

Discussion:

In this study we demonstrate that the vicilin LSs from PN and WN vicilins contain a variable number of VBP motifs, which adopt a characteristic a-hairpin fold mediated by disulfides between the conserved CxxxC elements. While similar structures have been described for other VBPs [14, 27, 33], this is the first time such a fold has been observed in major food allergens. This shared structure contains IgE binding sequences that were previously identified by their similarity to dominant epitopes of the 2S albumins, including one major epitope that lies in the loop between the paired disulfides. The structural and IgE binding data shown here adds more evidence to our previous studies [8, 9, 34] for the role of these LSs in mediating cross-reactivity between WN and PN, as IgE in sera from patients with dual allergy to both PN and WN showed particularly strong binding to this region. We also note that similarly located sequences from the LS of cashew (Ana o 1) and pistachio (Pis v 3) vicilins bind IgE from patients with the respective allergies and could contribute to the the clinically significant cross-reactivity between these two foods [35, 36].

While the sequences studied here have similar PCPs, they are quite different in their amino acid sequences. This suggests that the local structure of the area contributes to the recognition by the peptidases responsible for cleaving the pro region or LS from the mature vicilin. The α-hairpin structures are also quite pliable, as indicated by the few long range NOEs and the degree of disordered structure. This suggests that introducing mutations to remove specific epitopes should not greatly alter the structure. For example, while peptide A10, A11 and A12 contain areas around the inter-helical loop region with some of the disordered structure, A11 did not bind IgE from any of the tested sera. Thus, while the ordered hairpin structure can stabilize a variety of peptide sequences, the specific amino acid sequence are ultimately important for antibody binding.

Another consequence of the LS architecture is that its defining α-hairpin scaffold is maintained by only a few conserved interactions, allowing it to accumulate mutations over time without compromising its underlying structure. The resulting sequence diversity could allow LSs to mimic the surface features of unrelated proteins, giving rise to cross-reactivity between different allergen families as observed between J2LS and Ara h 2, as the most conserved region of Ara h 2 contains a similar pattern of cysteine residues separated by variable residues [7, 37].

The disparities in immunogenicity and cross-reactivity among VBPs, despite their shared structures, suggests there are additional factors involved. The differences in IgE binding to the J2LS fragments could be partially explained by examining J2LS trypsin digestion from a previous study which provides some insight into the fragments which may be formed upon gastric digestion [16]. Comparing this data with the NMR structures discussed above, we see that several arginine residues are protected from proteolysis due to their location within the structured CxxxC motif (Figure 1B), potentially contributing their downstream allergic potential. In sum, this illustrates that the mechanisms of cross-reactivity are complicated and involve both structural and amino acid sequence components.

Conclusions:

These results establish vicilin precursors as a family of evolutionarily distant, yet structurally similar proteins whose allergenic potential cannot be easily assessed using traditional sequence homology metrics. The cross-reactive peptides in the LSs are located in the most ordered region of the structures and are related in their PCPs to immunodominant epitopes of 2S-albumins. Concurrently, the widespread prevalence of what we refer to as repeat sequences, in a variety of common food sources, highlights the need for a combination of structural, bioinformatics and immunological tools to identify potential cross-reactivity between taxonomically distant allergen families. This emphasizes that while the α-hairpin structure is common and is likely to harbor important conformational epitopes, the location and sequence of the amino acids within the structure are likely to have a significant impact on the IgE binding affinity, the digestion sites, and the

clinical relevance of IgE cross-reactivity.

References:

1. Gupta, R.S., et al., *Prevalence and Severity of Food Allergies Among US Adults.* JAMA Netw Open, 2019. **2** (1): p. e185630.

2. Chong, K.W., et al., *Reaction phenotypes in IgE-mediated food allergy and anaphylaxis.* Annals of Allergy, Asthma & Immunology, 2020.**124** (5): p. 473-478.

3. Fischer, D., et al., *Anaphylaxis.* Allergy, Asthma & Clinical Immunology, 2018. **14** (2): p. 54.

4. Weinberger, T. and S. Sicherer, *Current perspectives on tree nut allergy: a review.* Journal of asthma and allergy, 2018.**11** : p. 41-51.

5. Maloney, J.M., et al., *The use of serum-specific IgE measurements for the diagnosis of peanut, tree nut, and seed allergy.* J Allergy Clin Immunol, 2008. **122** (1): p. 145-51.

6. García, B.E. and M.T. Lizaso, *Cross-reactivity syndromes in food allergy.* J Investig Allergol Clin Immunol, 2011. **21** (3): p. 162-70; quiz 2 p following 170.

7. Dreskin, S.C., et al., *The importance of the 2S albumins for allergenicity and cross-reactivity of peanuts, tree nuts, and sesame seeds.* Journal of Allergy and Clinical Immunology.

8. Maleki, S.J., et al., *Computationally predicted IgE epitopes of walnut allergens contribute to cross-reactivity with peanuts.* Allergy, 2011. **66** (12): p. 1522-1529.

9. Nesbit, J.B., et al., *Epitopes with similar physicochemical properties contribute to cross reactivity between peanut and tree nuts.*Mol Immunol, 2020. **122** : p. 223-231.

10. Shewry, P.R., J.A. Napier, and A.S. Tatham, *Seed storage proteins: structures and biosynthesis.* Plant Cell, 1995. **7** (7): p. 945-56.

11. Shewry, P.R., et al., *Plant protein families and their relationships to food allergy.* Biochem Soc Trans, 2002. **30** (Pt 6): p. 906-10.

12. Müntz, K., *Proteases and proteolytic cleavage of storage proteins in developing and germinating dicotyledonous seeds.* Journal of Experimental Botany, 1996. **47** (5): p. 605-622.

13. Lawrence, M.C., et al., *Structure of phaseolin at 2.2 A resolution. Implications for a common vicilin/legumin structure and the genetic engineering of seed storage proteins.* J Mol Biol, 1994.**238** (5): p. 748-76.

14. Zhang, J., et al., *An Ancient Peptide Family Buried within Vicilin Precursors.* ACS Chemical Biology, 2019. **14** (5): p. 979-993.

15. Yamada, K., et al., *Multiple functional proteins are produced by cleaving Asn-Gln bonds of a single precursor by vacuolar processing enzyme.* J Biol Chem, 1999. **274** (4): p. 2563-70.

16. Downs, M.L., et al., *Characterization of Low Molecular Weight Allergens from English Walnut (Juglans regia).* Journal of Agricultural and Food Chemistry, 2014. **62** (48): p. 11767-11775.

17. Aalberse, R.C., et al., *Identification of the amino-terminal fragment of Ara h 1 as a major target of the IgE-binding activity in the basic peanut protein fraction.* Clin Exp Allergy, 2020. **50** (3): p. 401-405.

18. Ivanciuc, O., et al., *The property distance index PD predicts peptides that cross-react with IgE antibodies.* Mol Immunol, 2009.**46** (5): p. 873-83.

19. Ivanciuc, O., et al., *Structural analysis of linear and conformational epitopes of allergens.* Regul Toxicol Pharmacol, 2009.**54** (3 Suppl): p. S11-9.

20. Midoro-Horiuti, T., et al., *Structural basis for epitope sharing between group 1 allergens of cedar pollen.* Mol Immunol, 2006.**43** (6): p. 509-18.

21. Lu, W., et al., *Distinguishing allergens from non-allergenic homologues using Physical-Chemical Property (PCP) motifs.* Mol Immunol, 2018. **99** : p. 1-8.

22. Fushman, D., et al., *Backbone dynamics of ribonuclease T1 and its complex with 2'GMP studied by two-dimensional heteronuclear NMR spectroscopy.* J Biomol NMR, 1994. **4** (1): p. 61-78.

23. Mueller, G.A., et al., *Backbone dynamics of the RNase H domain of HIV-1 reverse transcriptase.* Biochemistry, 2004. **43** (29): p. 9332-42.

24. Favier, A. and B. Brutscher, *Recovering lost magnetization: polarization enhancement in biomolecular NMR.* J Biomol NMR, 2011.**49** (1): p. 9-15.

25. Ghosh, D., et al., *Primary identification, biochemical characterization, and immunologic properties of the allergenic pollen cyclophilin cat R 1.* J Biol Chem, 2014. **289** (31): p. 21374-85.

26. Braun, B.A., C.H. Schein, and W. Braun, *Δ-γραπη γλυστερς φλαιρυσες ανδ β-γοροναιρυσες αγγορδινγ το τηειρ ηοστς, δισεασε τψπε ανδ ηυμαν γελλ ρεγεπτορ.* bioRxiv : the preprint server for biology, 2020: p. 2020.08.13.249649.

27. Oparin, P.B., et al., *Buckwheat trypsin inhibitor with helical hairpin structure belongs to a new family of plant defence peptides.*Biochem J, 2012. **446** (1): p. 69-77.

28. Payne, C.D., et al., *Defining the Familial Fold of the Vicilin-Buried Peptide Family.* J Nat Prod, 2020. **83** (10): p. 3030-3040.

29. Dall'Antonia, F. and W. Keller, *SPADE web service for prediction of allergen IgE epitopes.* Nucleic Acids Res, 2019.**47** (W1): p. W496-w501.

30. Burks, A.W., et al., *Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity.* Eur J Biochem, 1997.**245** (2): p. 334-9.

31. Cong, Y.J., et al., *Characterisation of the IgE-binding immunodominant epitopes on Ara h1.* Food and Agricultural Immunology, 2008. **19** (3): p. 175-185.

32. Beyer, K., et al., *Measurement of peptide-specific IgE as an additional tool in identifying patients with clinical reactivity to peanuts.* J Allergy Clin Immunol, 2003. **112** (1): p. 202-7.

33. Ng, Y.-M., et al., *Structural characterization and anti-HIV-1 activities of arginine/glutamate-rich polypeptide Luffin P1 from the seeds of sponge gourd (Luffa cylindrica).* Journal of Structural Biology, 2011. **174** (1): p. 164-172.

34. Nesbit, J.B., et al., *Identification and assessment of the IgE epitopes of Ara h 1 and Jug r 2 leader sequences.* Journal of Allergy and Clinical Immunology, 2018. **141** (2): p. AB179.

35. Willison, L.N., et al., *Pistachio vicilin, Pis v 3, is immunoglobulin E-reactive and cross-reacts with the homologous cashew allergen, Ana o 1.* Clin Exp Allergy, 2008. **38** (7): p. 1229-38.

36. Wang, F., et al., *Ana o 1, a cashew (Anacardium occidental) allergen of the vicilin seed storage protein family.* J Allergy Clin Immunol, 2002. **110** (1): p. 160-6.

37. Ivanciuc, O., et al., *Characteristic motifs for families of allergenic proteins.* Molecular Immunology, 2009. **46** (4): p. 559-568.

38. Ling, C., et al., *Expression and refolding of mite allergen pro-Der f1 from inclusion bodies in Escherichia coli.* Protein Expr Purif, 2015. **109** : p. 93-8.

39. Park, S.S., et al., *Primary structure and allergenic activity of trypsin inhibitors from the seeds of buckwheat (Fagopyrum esculentum Moench).* FEBS Lett, 1997. **400** (1): p. 103-7.

40. Micsonai, A., et al., *Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy.* Proc Natl Acad Sci U S A, 2015. **112** (24): p. E3095-103.

41. Micsonai, A., et al., *BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra.* Nucleic Acids Res, 2018. **46** (W1): p. W315-w322.

42. Hura, G.L., et al., *Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS).* Nat Methods, 2009. **6** (8): p. 606-12.

43. Dyer, K.N., et al., *High-throughput SAXS for the characterization of biomolecules in solution: a practical approach.*Methods Mol Biol, 2014. **1091** : p. 245-58.

44. Classen, S., et al., *Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source.* J Appl Crystallogr, 2013. **46** (Pt 1): p. 1-13.

45. Putnam, C.D., et al., *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution.* Q Rev Biophys, 2007. **40** (3): p. 191-285.

46. Farrow, N.A., et al., *Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation.* Biochemistry, 1994. **33** (19): p. 5984-6003.

47. Dall'Antonia, F., et al., *Prediction of IgE-binding epitopes by means of allergen surface comparison and correlation to cross-reactivity.* J Allergy Clin Immunol, 2011. **128** (4): p. 872-879.e8.

48. Shen, Y., et al., *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.* J Biomol NMR, 2009. **44** (4): p. 213-23.

49. Lee, W., et al., *PONDEROSA, an automated 3D-NOESY peak picking program, enables automated protein structure determination.*Bioinformatics, 2011. **27** (12): p. 1727-8.

50. Rosato, A., et al., *CASD-NMR: critical assessment of automated structure determination by NMR.* Nat Methods, 2009. **6** (9): p. 625-6.

51. Güntert, P. and L. Buchner, *Combined automated NOE assignment and structure calculation with CYANA.* J Biomol NMR, 2015.**62** (4): p. 453-71.

52. Bhattacharya, A., R. Tejero, and G.T. Montelione, *Evaluating protein structures determined by structural genomics consortia.*Proteins, 2007. **66** (4): p. 778-95.

53. Sánchez-Ruano, L., B. de la Hoz, and J. Martínez-Botas,*Clinical utility of microarray B-cell epitope mapping in food allergies: A systematic review.* Pediatr Allergy Immunol, 2020.**31** (2): p. 175-185.

Table 1. Patient allergies

Allergy	Patient	Age	C	A	PC	HZ	Other Nut	Other allergy
WN	1	54	[?]		[?]	[?]	pi, b	sesame
WN	2	33	[?]	[?]	[?]	[?]	pi, b	pumpkin
WN	3	22		[?]	[?]	[?]		coconut
WN	4	NA						
WN	5	50	[?]	[?]	[?]	[?]	pi, b	coconut
WN	6	38	[?]	[?]	[?]	[?]	pi, b	coconut, fish
WN	7	NA						
WN	8	53	[?]		[?]	[?]	pi, b	tomatoes
WN	9	NA						

Allergy	Patient	Age	C	A	PC	HZ	Other Nut	Other allergy
WN	10	48		[?]	[?]	[?]	m	apple, pear
WN	11	40	[?]	[?]	[?]	[?]	pi	sesame, banana, orange
WN	12	36		[?]	[?]			banana, cantaloupe
PN	13	NA						sunflower
PN	14	24						soy
PN	15	NA						
PN	16	35	[?]	[?]				soy
PN	17	12	[?]			[?]	pi, pn	
PN	18	40						
PN	19	25						
PN	20	NA						
PN	21	NA						
PN	22	NA						
PN	23	19						
PN	24	29						shrimp
PN/WN	25	63	[?]	[?]	[?]	[?]	pi, b	
PN/WN	26	25	[?]	[?]	[?]	[?]	pi, b	
PN/WN	27	21	[?]					
PN/WN	28	38		[?]	[?]			coconut
PN/WN	29	42	[?]	[?]	[?]	[?]	m, b	
PN/WN	30	48	[?]	[?]	[?]	[?]	pi, b	sesame
PN/WN	31	62	[?]	[?]	[?]	[?]	m, b	
PN/WN	32	38	[?]		[?]	[?]	m	shellfish
PN/WN	33	36	[?]	[?]	[?]		pn	shellfish
PN/WN	34	25	[?]		[?]			
PN/WN	35	36						
PN/WN	36	33						
PN/WN	37	17	[?]	[?]	[?]	[?]	pn, pi	shellfish, flax, soy
PN/WN	38	27	[?]	[?]			pn	poppy seeds
PN/WN	39	30						
PN/WN	40							sunflower

Acronyms: **WN** =walnut; PN=peanut; NA=not available; pi=pistachio; b=brazil nut; pn=pine nut, m=macadamia

Figure Legends:

Φιγυρε 1: ΠΝ ανδ ΩΝ ΄ΒΠαμεντς αδοπτ α ςομμον α-ηαιρπιν στρυςτυρε δεσπιτε λοω σεχυενςε ιδεντιτψ. A) NMR structures of J2LS and A1LS VBPs. Cartoon figures depict representative conformations of each structure. Inter-helix disulfides shown as sticks and highlighted in yellow. Complete NMR data tables are available in <blank>. B) Schematic view illustrating the disulfide bonding pattern for all three fragments. Trypsin cut sites identified by Downs et. al. are indicated by the yellow arrows. C) Identity matrix for the amino acid sequence of the VBPs.

Figure 2: Potential PN cross-reactive regions on A1LS. A) Sequence of A1LS used in the peptide microarray analysis. Residues depicted in the solution-NMR structures are shown in capital letters. Residues from the expression system used to generate the NMR samples but are not present in the peptide microarrays are denoted with brackets. Residues highlighted in blue represent alpha-helices identified in the available NMR structures. Epitopes identified by Burks et. al are indicated in grey rectangles, whereas cross-reactive peptides identified are indicated by coloured rectangles. AEP cut site is indicated by the grey arrow. B) Structure of A1LS coloured by surface similarity (dSIM) values against J2.1, J2.2, and J2.3 as calculated

using the SPADE surface comparison algorithm. Total dSIM ($\Sigma[\Delta SIM]$) for all residues is indicated. C) IgE-reactive peptides mapped onto the structure of A1LS. Peptides colour-coded as in (A).

Figure 3: Potential PN cross-reactive regions on J2LS.A) Structure of J2.1, J2.2, and J2.3 coloured by surface similarity (dSIM) values against A1LS as calculated using the SPADE surface comparison algorithm. Total dSIM ($\Sigma[\Delta SIM]$) for all residues is indicated. B) Sequence of J2LS, showing the three repeat areas (Nesbit 2020), with helical areas highlighted in blue, residues used to obtain the 31 mer consensus sequence underlined, and the central 13mer consensus in bold letters. Residues included in the solution-NMR structures are capitalized; those included in the NMRsamples but not present in the peptide microarrays are denoted in parentheses. Epitopes identified by Maleki et. al 2011 are indicated in grey rectangles, with cross-reactive peptides identified in this study (Fig. 4B) shown in coloured rectangles labeled JR5-JR9. The AEP cut site is indicated by the grey arrow. C) IgE-reactive peptides mapped onto the structure of J2.1, J2.2, and J2.3. Peptides colour-coded as in (A). For the sake of simplicity, only JR6-JR9 are shown.

Figure 4: Comparison of IgE binding to the A1LS and J2LS among walnut (WN), peanut (PN), and dual allergic (PW) patients by microarray analysis . A) Modified z values were calculated from the microarray median signal-to-noise ratios (SNRs) and were used to determine the percentage of IgE binding within an amino acid range (shown in 2^{nd} column) for each allergy group. B) The percentage of patients which bind IgE to peptides and major epitopes of A1LS and J2.1 and J2.2 is shown. A major epitope was identified as a peptide where 50% or more of the P, W or PW patients exhibited a positive binding event (z value of 3 or greater) for that peptide. The epitopes marked with an asterisk were recognized by all three allergy groups. The shared regions of overlapping peptides are underlined. Average z values are given as low (3 – 6), medium (6.1 – 9) and high (>9 and in red text). Amino acid composition is shown as percentage of hydrophobic (H), acidic (A), basic (B) or neutral (N) amino acids in the peptide sequence.

Figure 5: Property distance (PD)-graph showing the interrelatedness of the IgE binding epitopes of the LS to those of the 2S albumins of peanuts and tree nuts. The sequences of the LS peptides identified as IgE binding in this study (Figure 4B) are similar in their similar PCPs to the consensus and previously identified epitopes from PN and TN 2S albumins. The PD graph is the graphical depiction of the most similar sequences, which are connected by blue lines, and the lines descrease in intensity and width with increasing PD (lower similarity). Note that the previously determined PCP-consensus sequences (Nesbit et al. 2020) lie in the middle of the graph, and that the LS IgE reactive peptides show similarity to the 2S albumins.

Supplemental Material Methods:

Constructs and Purification

The A1LS has been previously observed in peanut extracts[17]. The sequence of the observed peptide with minor modifications was used to generate the sequence referred to as A1LS in this paper, but whose formal designation is Ara h 1.0101 (25-83) A25G. Constructs for JR2.1: Jug r 2.0101 (1-57) M0 added , JR2.2: Jug r 2.0101 (67-111) Q67G N68M, , and JR2.3: Jug r 2.0101 116-161 D116D R117M were prepared by identifying their constituent CxxxC pairs and extending the sequence 10-20 residues on either side. All constructs were cloned into the pDest expression system with an N-terminal Glutathione S-transferase (GST) affinity tag separated from the main sequence by a tobacco etch virus (TEV) protease cleavage site. Cells expressing these constructs were grown in 2x YT media to an OD of ˜0.8 at 37°C. To produce uniformly ^{15}N-labeled or^{13}C-^{15}N-labeled A1LS for NMR, cells were grown overnight in 1 L Luria Broth (LB), harvested, and subsequently transferred to M9 media with$^{15}NH_4Cl$ and, where applicable,^{13}C-glucose as the sole nitrogen and carbon sources respectively. Expression was induced with 0.5 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) at 16°C overnight. A1LS was purified from the resulting cells using an immobilized glutathione column, eluted with 10 mM reduced glutathione in pH 7.4 PBS, and incubated with TEV protease overnight to cleave the GST tag. Correct disulfide bond pairing was achieved by diluting the sample and adding oxidized glutathione to a final concentration of 2 mM, and 0.5 mM reduced and oxidized glutathione respectively [38] (redox buffer), the efficacy of which was subsequently verified using^{1}H-^{15}N NMR and mass spec (S2),

and further confirmed over the course of 3D structure determination described subsequently. Following 30 minute incubation the sample was loaded onto a Superdex75 26/600, and eluted with PBS to remove the GST tag. The J2LS fragments J2.1, J2.2, and J2.3 were purified in a similar manner with the following modification: the GST tag was removed via incubation with TEV protease in the presence of 2 mM DTT, the latter of which was required to ensure complete cleavage for some constructs. The buffers in the resulting samples were exchanged for PBS, incubated with redox buffer, and purified via SEC as described above.

The full length (FL) J2LS as identified by Downs *et al.* [16] but with the addition of an N-terminal start codon (Met) and without the C-terminal Arg was cloned into the Pet9a vector. The protein was expressed in BL21 cells as above. Cells were lysed and the resulting soluble proteins were precipitated using 25%, 50%, and 75% ammonium sulfate. The 75% precipitate fraction was isolated and resuspended in Buffer A (50 mM Tris-HCL pH 8.3, 100 mM NaCl) and loaded onto a Mono-S-Sephrose anion exchange chromatography column. FL JR2 was eluted using a linear gradient of Buffer A and Buffer B (50 mM Tris-HCL pH 8.3, 500 mM NaCl). FL J2LS-containing fractions were incubated for 30 minutes with 2 mM/0.5 mM of reduced/oxidized gluthathione to ensure proper disulfide bond formation before being loaded onto a Superdex75 26/600 sizing column and eluted with PBS to yield the final purified protein.

Biophysical Characterization

Circular dichroism spectra were collected on 0.5 µM Fl J2LS, or 2 mM A1LS, J2.1, J2.2, or J2.3 in PBS at 25°C using a Jasco J-815 CD spectropolarimeter. Each spectrum represents the average of four accumulations scanning at 20 nm/ min with a data-pitch of 0.2 nm. Each spectrum was acquired from at least three trials representing two biological replicates unless specified. Secondary structure prediction was carried out using the BESTSEL web server [39, 40].

Samples of 46 uM or 23 uM FL J2LS in PBS were prepared and analyzed using the SIBYL High-Throughput mail-in service. The resulting data was analyzed using the ScÅtter software [40-45].

NMR

Triple-resonance and NOESY spectra were collected on 0.1-1 mM protein samples in PBS using either a 600 or 800 MHz Varian DD2 console equipped with cryogenically cooled probe. Backbone and side-chain assignments, and NOESY distance restraints were obtained using standard triple resonance techniques employing either the standard VARIAN Biopack or modified BEST-TROSY pulse sequences [24, 25]. T_1 and T_2 relaxation times were obtained for the ^{1}H-^{15}N peaks using a modified HSQC pulse sequence, and used to determine T_c as described previously [46].

Epitope prediction and ΔSIM values were obtained using the SPADE (Surface comparison-based Prediction of Allergen Discontinuous Epitopes) algorithm access via the provided web server. NMR structures A1Pro and the J2LS fragments were used as inputs. Predicted cross-reactivity was listed as 55% between A1LS and the J2LS fragments, or 100% within the JR22 fragment. These values were selected to broadly reflect the cross-reactive IgE binding observed in the microarray studies[29, 47].

NMR Structure Calculation:

Backbone and side-chain assignments were obtained for over 90% of the residues for A1LS and the J2LS fragments using standard triple-resonance experiments. Secondary structure was estimated from the available chemical shifts using TALOS and found to be consistent with the proposed structures (S2, S3) [48]. Cystine CB chemical shifts were uniformly >33 ppm in the oxidized form suggesting that they were all involved in disulfide bonds, providing further verification of the final structures (S4).

NOESY spectra in conjunction with the backbone and side chain assignments were used to calculate the 3D structure via the PONDEROSA server[49, 50]. All the LS fragments has a paucity of true long-range NOEs constraints, even relative to other vicilin precursor fragments (S6). As such, multiple iterations of CYANA were employed to identify putative long-range NOEs [51], which were verified manually and resubmitted to the PONDEROSA for further refinement until no new putative long-range NOEs were suggested by the

software. NMR structural statistics were calculated using the Protein Structure Validation Software (PSVS) suite are shown in S6 [52].

Statistical analysis of microarray data

Prior to analysis, we collected the median signal to noise ratio (SNR) for each peptide spot contained in the J2LS and A1LS for each patient. We chose SNR as our signal measurement as it corrects the raw signal intensity for non-specific hybridization and instrument noise (link). Each SNR value was then converted into modified z-scores which are calculated using median and median absolute deviation (MAD) rather than mean and standard deviation. Z-scores represent a standardized signal intensity and are frequently used to report immunoglobulin binding to peptide microarray [53]. Median and MAD were calculated for each patient and leader sequence combination. MAD was calculated with the*mad* function in R using the constant of 1.4826 to approximate standard deviation. After calculating patient and LS-specific median and MAD, z-scores were calculated by subtracting the median from a relevant spot SNR and dividing this value by the MAD. We defined a true IgE binding event as a SNR with a converted Z score >= 3.

Statistical analyses were performed using R (version 3.6.3; R Development Core Team, available at www.r-project.org). A goal of this study was to assess cross-reactivity among walnut-, peanut- or co-allergic individuals to regions of A1LS and J2LS. For each LS and predicted J2LS fragment, we counted the number of significant IgE binding events (z-value > 3) to their representative microarray peptides for each allergy type: walnut (WN), peanut (PN), or co-allergy (PW). For each LS or J2LS fragment we determined the percent of peptides bound by IgE for each allergy group. For example, JR2.1 is represented by 11 peptides on each microarray and there are 12 total walnut allergic individuals meaning that there is a total of 132 possible IgE binding events. As 42 of these peptides exhibited IgE binding with a z-value >3, we determined that walnut-allergic patients bound 32% of the possible JR2.1 microarray peptides. We then used the *fisher.test* function to perform Fisher's exact tests to compare the percent of peptides with significant IgE binding among the allergy groups for each LS or J2LS fragment.

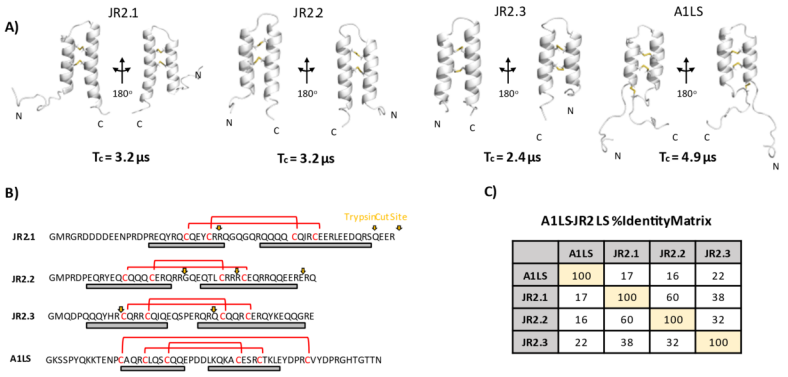
Supplemental Material Figure legends

S1: Initial characterization of A1LS and JR2LS. A) ^{1}H-^{15}N HSQC of A1LS before and after redox buffer. Magenta contours show purified A1LS in PBS buffer. The blue contours show A1LS following incubation with reduced and oxidized glutathione (2 and 0.5 mM respectively). New peaks following treatment are labeled. B) TALOS secondary structure prediction for A1LS based on backbone and side-chain chemical shifts. The two helical regions of the predicted α-hairpin fold are clearly identifiable. C) ^{1}H-^{15}N-NMR spectrum of FL JR2 LS displaying numerous well-dispersed peaks consistent with an alpha-helical protein. D) Circular dichroism spectra of indicated hairpinin and the table below showing the percentage of a-helices, b-sheets and random coils within each.

S2: Additional NMR characterization of J2LS fragments. Predicted secondary structure of JR2 VBPs as calculated based on the backbone and side chain NMR chemical shifts using the TALOS algorithm.

S3: Additional NMR data . A) Data table showing various parameters relating to the NMR structure determination and validation of A1pro and JR2 LS fragments. B) Cysteine CB chemical shift values for all four constructs. Values indicate that all cystines are involved in disulfide bonds.

S4: Surface comparison with distant VBP's. A) NMR structure of VBP-8; and B) BWI-2C VBP's from tomato and buckwheat respectively. Spade surface-similarity comparison with immunodominant A1LS and JR2.1. shown below, revealing areas of high surface similarity despite the lack of known tomato-nut and buckwheat-nut cross-reactivity . PDB ID: 6O3S, 2LQX



	A1LS	JR2.1	JR2.2	JR2.3
A1LS	100	17	16	22
JR2.1	17	100	60	38
JR2.2	16	60	100	32
JR2.3	22	38	32	100

Figure 1

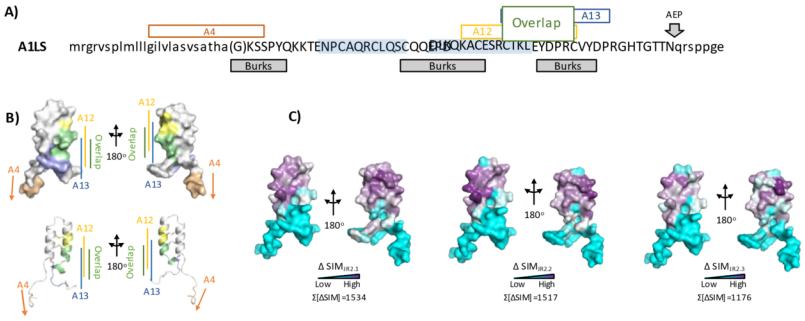


Figure 2

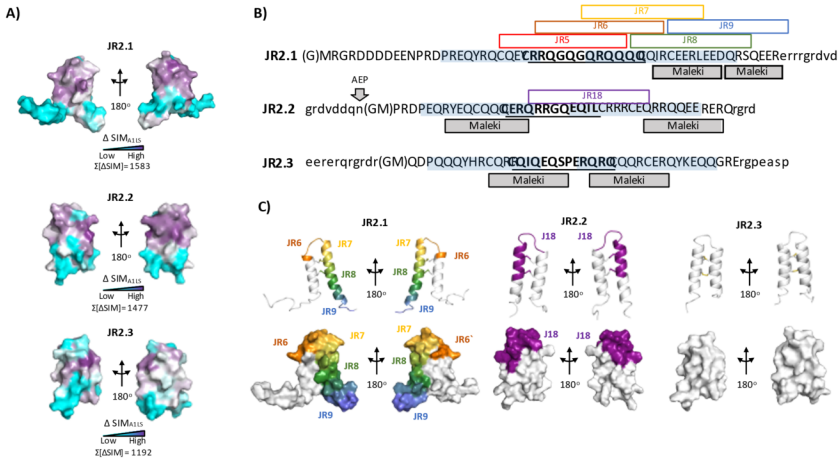


Figure3

A)

Protein	Amino acid range	% of microarray peptides with significant IgE binding (z >3) within the amino acid range				
		WN	PN	PW	P-value	Fisher's test results
J2LS	1 - 175	14	23	20	0.005	PN>WN, PW>WN, PN=PW
JR2.1	1 - 65	32	49	43	0.014	PN>WN, PW>WN, PN=PW
JR2.2	60 - 115	6	14	11	0.192	N/D
JR2.3	106 - 175	2	3	6	0.224	N/D
A1LS	1 - 90	8	17	13	0.028	PN>WN, PW>WN, PN=PW

B)

Peptide #	Amino acid sequence	% IgE binding			Average z value			Amino acid % composition			
		WN	PN	PW	WN	PN	PW	H	A	B	N
A4	VLASVSATHAKSSPY	25	58	37.5	2.9	3.5	2.2	47	0	13	40
A10	CQQEPDDLKQKACES	8	42	19	1	5.1	6.4	20	27	13	40
A11	DDLKQKACESRCTKL	0	0	0	0.7	0.4	0.5	20	20	27	33
A12	KACESRCTKLEYDPR	42	50	50	3.1	4.2	3.2	20	20	27	33
A13*	RCTKLEYDPRCVYDP	50	92	81	5.3	8.1	7.2	27	20	20	33
JR4	EQYRQCQEYCRRQGQ	0	33	19	0.6	2.85	2.65	0	13	20	67
JR5	CQEYCRRQGQGQRQQ	17	58	50	2.43	4.82	3.99	0	7	20	73
JR6*	RRQGQGQRQQQQCQI	50	75	75	6.49	17.7	16.3	7	0	20	73
JR7*	GQRQQQQCQRCEER	75	100	75	10.9	17.2	16.3	7	13	20	60
JR8*	QQCQIRCEERLEEDQ	67	75	81	7.4	9.5	11.2	13	33	13	40
JR9*	RCEERLEEDQRSQEE	50	100	69	6.2	10.3	9.5	7	47	20	30
JR18*	QRRGQEQTLCRRRCE	50	83	69	4.8	8.8	6.8	7	13	33	47

Arbitrary cut off values for Z scores: Low=3-6; Med=6.1-9; High=>9WN=walnut allergic; PN=peanut allergic; PW=peanut and walnut allergic; Amino acid composition: H=hydrophobic; A=acidic, B=basic, N=Neutral

Figure4

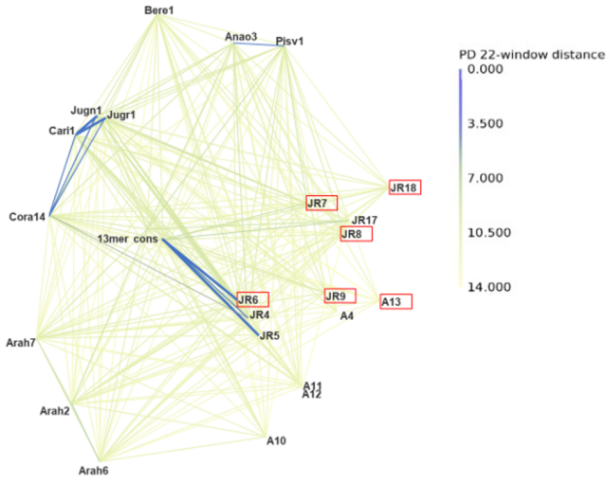


Figure5

Hosted file

LS final figures.pptx available at https://authorea.com/users/421401/articles/527458-structure-and-ige-cross-reactivity-among-walnut-and-peanut-vicilin-leader-sequences