



Epitopes with similar physicochemical properties contribute to cross reactivity between peanut and tree nuts

Jacqueline B. Nesbit^a, Catherine H. Schein^{c,*}, Benjamin A. Braun^b, Stephen A.Y. Gipson^a, Hsiaopo Cheng^a, Barry K. Hurlburt^a, Soheila J. Maleki^{a,*}

^a Dept of Agriculture-Agricultural Research Service-Southern Regional Research Center (USDA-ARS-SRRC), New Orleans, LA, United States

^b Department of Computer Science, Stanford University, United States

^c Department of Biochemistry and Molecular Biology, Institute for Human Infection and Immunity, University of Texas Medical Branch at Galveston (UTMB), TX, United States

ARTICLE INFO

Keywords:

Walnut allergy
Peanut allergy
Tree nut allergy
Cross reactivity
Structural Database of Allergic Proteins
Food allergy
Computational prediction
Epitopes
Allergen
IgE

ABSTRACT

Many individuals with peanut (PN) allergy have severe reactions to tree nuts (TN) such as walnuts or cashews. Although allergenic proteins in TN and PN have overall low identity, they share discrete sequences similar in physicochemical properties (PCP) to known IgE epitopes. Here, PCP-consensus peptides (cp, 13 aa and 31 aa) were identified from an alignment of epitope rich regions of walnut vicilin, Jug r 2, leader sequence (J2LS) and cross-reactive epitopes in the 2S albumins of peanut and synthesized. A peptide similarity search in the Structural Database of Allergic Proteins (SDAP) revealed a network of peptides similar (low property distance, PD) to the 13 aa cp (13cp) in many different plant allergens. Peptides similar to the 13cp in PN and TN allergens bound IgE from sera of patients allergic to PN and TN in peptide microarray analysis. The 13cp was used to produce a rabbit consensus peptide antibody (cpAB) that detected proteins containing repeats similar to the 13cp in western blots of various nut extracts, in which reactive proteins were identified by mass spectrometry. The cpAB bound more specifically to allergens and nut extracts containing multiple repeats similar to the 13 cp, such as almond (Pru du 6), peanut (Ara h 2) and walnut (Jug r 2). IgE binding to various nut extracts is inhibited by recombinant J2LS sequence and synthetic 31cp. Thus, several repeated sequences similar to the 13cp are bound by IgE. Multiple similar repeats in several allergens could account for reaction severity and clinically relevant cross-reactivity to PN and TN. These findings may help improve detection, diagnostic, and therapeutic tools.

1. Introduction

Up to 30 % of individuals with a clinically significant peanut (PN) allergy are also clinically allergic (not just sensitized) to at least one tree nut (TN) species (Bublin and Breiteneder, 2014; de Leon et al., 2007, 2003; Glaspole et al., 2011; Kulis et al., 2009; Maloney et al., 2008; Sicherer et al., 2003; Uotila et al., 2018). While the major TN and PN allergens can be grouped into similar protein families (PFAM) (Ivanciuc et al., 2009a; Schein et al., 2010), the overall sequence identity even within the same PFAM is quite low. For example, the peanut 2S albumin allergen, Ara h 2, is < 35 % identical to the related walnut allergen, Jug r 1 (Smeekens et al., 2018). This suggests their allergenicity lies in discrete sequence areas. Previously, the peptide similarity search program, in the Structural Database of Allergic Proteins (SDAP) (Ivanciuc

et al., 2009a, a; Ivanciuc et al., 2009b, 2002) was used to identify sequences with similar physicochemical properties (PCPs) to an immunodominant epitope of the 2S albumin, Ara h 2 and predict potential epitopes in other allergens (Maleki et al., 2011, 2012). The PD search revealed similar IgE epitopes in the leader sequence (LS) of the walnut allergen, the 7S globulin Jug r 2 (J2LS). These repeated sequences were recognized by serum IgE of both walnut and peanut allergic individuals. Therefore the walnut allergen Jug r 2, whose mature region has a vicilin fold similar to Ara h 1, has several IgE reactive sequences in its LS that are similar (low property distance, PD) to an immunodominant epitope of Ara h 2, a 2S albumin. Here we address the question of whether repeats with similar PCPs to a consensus sequence could play a role in IgE binding, and cross reactivity of peanuts and nuts.

Plant proteins are known to contain repeated sequences, due to the

* Corresponding author at: U.S. Dept of Agriculture-Agricultural Research Service-Southern Regional Research Center (USDA-ARS-SRRC), 1100 Robert E. Lee Blvd, New Orleans, LA, 70124, United States.

** Corresponding author at: UTMB, 301 University Blvd., Galveston, TX, 77555, United States.

E-mail addresses: chschein@utmb.edu (C.H. Schein), soheila.maleki@usda.gov (S.J. Maleki).

<https://doi.org/10.1016/j.molimm.2020.03.017>

Received 16 October 2019; Received in revised form 11 February 2020; Accepted 26 March 2020

Available online 19 May 2020

0161-5890/ Published by Elsevier Ltd.

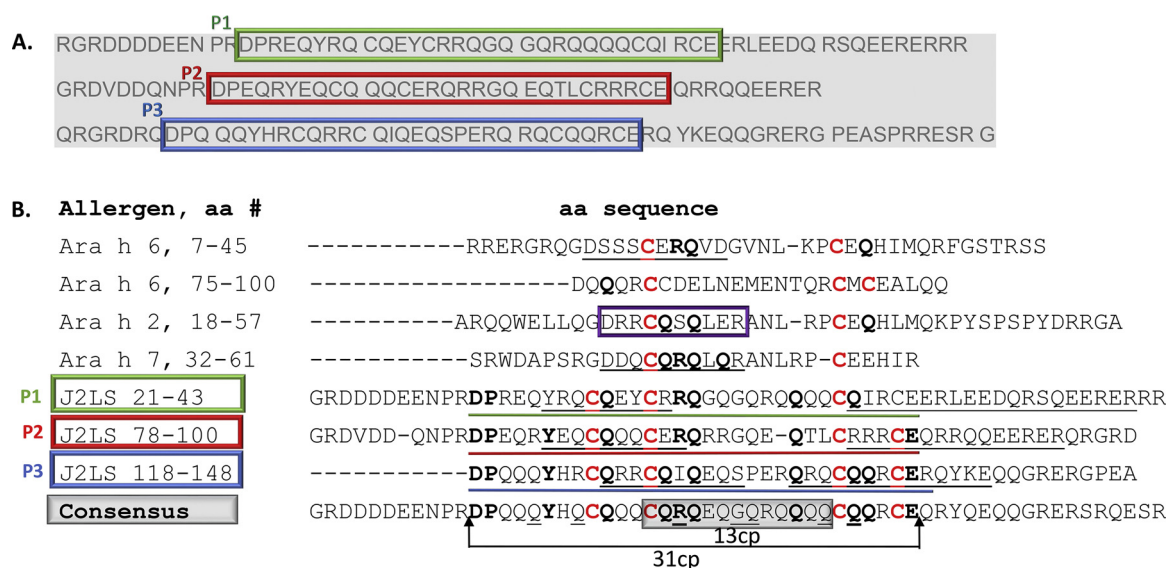


Fig. 1. Derivation of a PCP-consensus of epitope regions of 2S albumins and the J2LS. **A.** The J2 LS contains 3 cysteine-rich repeat sequences shown in red, green and blue. **B.** These can be aligned with similar, epitope rich regions of the 2S albumins, Ara h 2,6,7. The black, underlined regions correspond to the peptide sequences with low PD to the immunodominant epitope of Ara h 2, DRRCQSQLER (purple box). The 31aa consensus peptide (31cp) sequence (DPQQQYHQCCQCCQRQEQGQRQQCCQQRCE) indicated by the arrows. Underlined amino acid (aa) indicates that it is shared with at least one other J2LS region, conserved aas in the alignment are in bold, and cysteines (C) are in red. The grey, boxed region on the last line of sequences, labeled Consensus (in left column), is the 13cp sequence (CQRQEQGQRQQCCQ) used to generate cpAB and to perform PD searches in SDAP.

action of transposons and the polyploidy of their genomes. For example, repeated sequences account for as much as 67 % of cotton genomes (Li et al., 2015) and microsatellite repeats are used to classify cultivars of plants such as peanuts (Wang et al., 2016), rice (Kaur et al., 2015) flowering cabbage (Chen et al., 2017) and many others. Plant selection can also contribute to gene duplications if the sequences enhance characteristics such as stability to temperature fluctuation, growth rate or resistance to environmental stresses. One would expect that the genes of rapidly growing, short lived plants such as peanuts would evolve at a faster rate than the genome of TN (Luo et al., 2015) and therefore contain more repeats.

To determine whether other plant allergens contain peptides with similar potential epitopes (or repeats), the corresponding areas in peanut 2S albumin allergens (Ara h 2, h 6, and h 7) were aligned with the three long repeats of the J2LS (Fig. 1). The PCPcon program (Danecek et al., 2010; Danecek and Schein, 2010; Schein et al., 2012) was then used to calculate a PCP-consensus peptide (cp) for a 31 consensus peptide (31cp) and a shorter 13cp from the most conserved region of the trimeric repeat. Extending the search for peptides similar to the 13cp in SDAP revealed a network of similar peptide repeats in many other allergens, from different protein families. Further, rabbit polyclonal antibodies to the 13cp (cpAB) detected allergens predicted to contain repeated sequences in nut extracts. We empirically determined the binding of a specific antibody and IgE from patient sera to similar peptides from PN and TN allergens synthesized on microarrays. We found that the most reactive peptides were indeed those with the highest similarity (lowest PD) (Ivanciuc et al., 2004) to the 13cp.

The high similarity of the cross-reactive, repeated sequences in allergens from strikingly different plants and trees indicates they must have arisen from an ancient common ancestor. These results provide evidence that the 13cp represents an evolutionarily conserved region that may account for clinically significant cross-reactivity. This knowledge may be useful in anticipating proteins with high allergenic potential and improving diagnosis of cross-reactivity.

2. Methods

2.1. Patient sera

Sera from individuals allergic to PN and TN were collected at the Clinical and Translational Unit (CTU) of Tulane University School of Medicine, New Orleans, LA in accordance with the rules and regulations of the institutional review board of this institution and in accordance with US federal policy for the protection of human subjects. The patients are mostly over the age of 18, who have experienced recurrent severe, systemic, allergic reactions to peanut, walnut, cashew or almond. Adult sera was used, where their allergy is unlikely to change, involving relevant IgE epitopes. The patients were referred to the study based on their convincing history of food allergy and did not undergo food challenge for research purposes due to the potential severity of reactions Table 1 summarizes the allergic characteristics of the patient sera used in this study.

2.2. Microarray analysis

Synthetic overlapping 15 mer peptides, offset by five amino acids, covering the amino acid sequence of selected PN and TN allergens, were synthesized and printed onto microarray slides by JPT Peptide Technologies (JPT Peptide Technologies GmbH, Berlin, Germany). The slides were placed into a HS400 Pro TM (Tecan, San Jose, CA) where they were blocked in filtered Superblock (Thermo, Rockford, IL) for 30 min, at room temperature (RT), under agitation. The slides were washed with Tris-buffered saline containing 0.5 % Tween -20 (TBST). After centrifugation, 200 µl of patients' sera were injected into individual chambers containing the slides and incubated at 4 °C overnight (16 h) with agitation. The slides were washed and injected with 170 µl of mouse anti-human IgE (Life Technologies, Grand Island, NY), diluted in filtered Superblock, for 30 min at RT. The slides were washed, dried and scanned on a GenePix-4000B. IgE-binding was measured by the Cy3, green fluorescence at 532 nm.

Table 1
Patient allergies.

Allergy	Age/ Sex	PN	W	C	A	Other Nut	Other allergy
PN	36/M	✓					coconut, strawberry
PN	14/M	✓					
PN	32/F	✓					soy
PN	44/F	✓					pumpkin seed, shellfish, fish, eggplant
PN	35/F	✓					shellfish
PN	36/F	✓					shrimp*
PN	30/M	✓*					
PN	33/F	✓					
TN	35/F			✓		po, pi	mango
TN	56/M		✓		✓	pc, b, pi, h, m	
TN	58/M			✓		pi, m	shellfish
TN	35/F		✓		✓	pc, b, pi	fruit
TN	47/F		✓		✓	pc	fruit, carrots, sesame
TN	30/F			✓		*	
PN/TN	41/F	✓	✓	✓	✓		
PN/TN	26/F	✓				all	mango, peach, plum
PN/TN	25/M	✓	✓	✓	✓	all	shellfish, soy, flax
PN/TN	34/F	✓*		✓	✓	pi	sesame, poppy, coconut, berries
PN/TN	44/M	✓	✓	✓		pc, po, b, pi, h, m	shellfish
PN/TN	45/F	✓		✓			Shrimp, dairy
PN/TN	24/F	✓	✓	✓		pc	
PN/TN	32/F	✓		✓		pi, h, po	
PN/TN	54/M	✓	✓	✓	✓		
PN/TN	47/F	✓				TN, avoids all	
PN/TN	35/M	✓	✓		✓		soy
PN/TN	35/M	✓	✓	✓	✓	pc, pi	shellfish

Symbols: * = anaphylactic.

Acronyms: PN = peanut; W = walnut, C = cashew, A = Almond, pc = pecan, po = pistachio, pi = pine nut, h = hazelnut, m = macadamia, b = brazil nut.

2.3. Computational tools

The peptide similarity (PD search) (Ivanciuc et al., 2009b; Oezguen et al., 2008; Schein et al., 2005a) is one of several validated computational methods developed to enable searches within SDAP to identify sequences similar to known IgE epitopes (Ivanciuc et al., 2003a, 2004; Ivanciuc et al., 2002, 2003b). PD values < 8–10, depending on peptide length and overall composition, are statistically significant and have been shown to correlate with immune recognition (Maleki et al., 2011). Additional statistical measures (z-scores) are calculated to indicate the significance of the match relative to a random match (Ivanciuc et al., 2009b; Schein et al., 2005a, b). Peptides with identical sequences have a PD value of 0. In our validation tests, cross-reacting peptides generally have PD values below 9, although peptides with recognizable identity may occur up to about PD of 10 when longer search peptides are used (where the significance will be reflected in higher z scores).

2.3.1. Property distance graph (PD-graph)

Property distance graph (PD-graph) calculates the interpeptide PD for a group of sequences and graphically presents their similarity to one another at the PCP level (Braun, B.A. et al. submitted). 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 peptides. 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.

2.3.2. The PCPcon program

The PCPcon program, which was developed to define rational reference sequences for flaviviruses that would avoid numerical bias (Danecek et al., 2010; Danecek and Schein, 2010), was used to determine a PCP-consensus sequence of the three repeat areas of the J2LS,

as further described below. PD-graphs are alignment free and depend only on the linear sequence, without regard to structure. These factors may be better accounted for in PCPcon, whereby the alignment should be manually edited to remove fragments and insure that conserved residues, especially cysteines, are within the same column. Validation tests have shown that PCP-consensus sequences are closest in value (and % identity) to all members of the sequence set used to derive them, thus representing a true mean. A PCP-consensus envelope protein domain bound antibodies to all four Dengue virus serotypes (Bowen et al., 2012) and a PCP-consensus VPg was recognized by three diverse enterovirus polymerases (Schein et al., 2015).

2.4. Peptides and antibody

A PCP-consensus was calculated from the similar cysteine-rich repeats of the J2LS (Fig. 1). A 31cp (DPQQQYHQCCQQCQRQEQGQR-QQQCCQQRCE) was synthesized (Yale Peptide Synthesis Facility) for western blotting and competitive inhibition ELISA (ciELISA). A 13 aa central region (13cp = CQRQEQGQRQQQ), was used to make an anti-consensus sequence antibody in rabbits (cpAB) (Immunotech, Ramona, CA).

2.5. Purified proteins

Jug r 1-Walnut protein was extracted from walnut flour with a 20 mM Tris, 100 mM NaCl buffer at pH 8.5. This extract was then subjected to ammonium sulfate fractionation, followed by standard gravity flow, chromatography with Mono-Q High prep resin (BioRad Laboratories, Hercules, CA). The 70 % ammonium sulfate pellet was dissolved and loaded on the column in a 20 mM Tris, 50 mM NaCl buffer at pH 8.5 and eluted with a salt gradient of 50 mM–400 mM NaCl. Fractions containing of purified Jug r 1 was used here.

Jug r 2LS-The cDNA clone for Jug r 2 LS was synthesized by EZbiolab (EZbiolab Inc., Parsippany, NJ), cloned into pET-9A expression vector and expressed in BL21 E. coli. The soluble protein was extracted from bacterial pellet by standard means and subjected to ammonium sulfate precipitation. The 75 % ammonium sulfate pellet was dissolved in a 50 mM Tris, 100 mM NaCl, 1 mM EDTA buffer at pH 7 and loaded onto a Mono-S High prep resin (BioRad Laboratories, Hercules, CA) and eluted with a salt gradient of 100 mM–400 mM NaCl. Fractions containing of purified Jug r 2LS were used here.

2.6. SDS-PAGE and western blot

All nut flours were purchased from Oh! Nuts (www.ohnuts.com) as nut meals and defatted in our laboratory by the Soxhlet standard mode of a Buchi-B-811 extraction unit (Buchi Corp., New Castle, DE) for 2 h for each 10 g of nut flour. Defatted nut flour samples were mixed with 3x sample loading buffer (Invitrogen Corp., Carlsbad, California), incubated for 10 min at 65 °C, subjected to SDS-PAGE (Invitrogen Corp.), stained with Gel-Code Blue stain (Pierce, Rockford, IL), destained and photographed. Following SDS-PAGE, the samples were transferred to a PVDF membrane, which was blocked with 2 % nonfat dried milk dissolved in phosphate buffered saline containing 0.5 % Tween 20 (PBST) and incubated for one hour with a 1:5000 dilution of cpAB, washed and incubated for 30 min with anti-rabbit IgG (1:100,000) horseradish peroxidase (HRP)-labeled secondary antibody (Sigma Chemical Company, St. Louis, MO) diluted in blocking buffer. The membranes were washed and incubated with ECL-Plus Western substrate (Amersham Bioscience Corp., Piscataway, NJ). The signal was visualized using a CCD camera system (Fuji Photo Film Co., Ltd., Duluth, GA). Magic Mark (MM), molecular weight marker (Invitrogen Corporation, Carlsbad, California) was used.

A. Many allergens contain one or more sequences similar (low PD value) to the 13cp.

Allergen	Source	# of repeats	PD Value	Predicted cross-reactive sequences
Pru du 6	Almond	>10	4.95	QQEQLQQERQQQ
Jug r 2LS	Walnut	>7	5.45	CRRQGQGRQQQ
Ber e 1	Brazil nut	2	5.85	QQCREQMERQQQ
Cor a 9	Hazelnut	5	5.97	WERQERQERESE
Car i 4	Pecan	5	6.31	QQEYEQHRRQQQ
Jug r 4	Walnut	4	6.31	QQEYEQHRRQQQ
Pis v 2	Pistachio	7	6.39	SQRKSQSERERQ
Vig r 2	Mung bean	5	7.27	QQQQQQQESQQE
Pis v 3	Pistachio	2	7.35	VQRQYDEEKEQ
Ara h 4	Peanut	2	7.44	DQSQQQDSDSHK
Ses i 6	Sesame	2	7.47	MERTEASEQQDR
Sin a 2	Mustard	1	7.53	GSSQQQQQQQQN
Pis v 5	Pistachio	1	7.91	QQQQQQQSRGR
Bra i 1	Turnip	2	7.92	VRQQGQQQMQQ
Ric c 1	Ricin	3	7.94	NQQQESQLQQC
Ana o 1	Cashew	3	8.02	CQERYKKERQGH
Ses i 3	Sesame	4	8.12	ERRREEKYRQQ
Ara h 7	Peanut	1	8.2	MRRRVEQEQQE
Ara h 1	Peanut	3	8.23	AVRKEQQQRGR
Ara h 2	Peanut	1	8.27	QSDRLQGRRQQEQ
Ara h 3	Peanut	7	8.28	YQQQSRQSRRRS
Cor a 11	Hazelnut	2	8.4	CRDERQFDEQQR
Ana o 2	Cashew	1	8.5	QQQQQHQSRRGN
Cor a 14	Hazelnut	2	8.8	SNQQQQQELEQC
Ana o 3	Cashew	3	8.9	CQNLEQMVRQLQ

B.

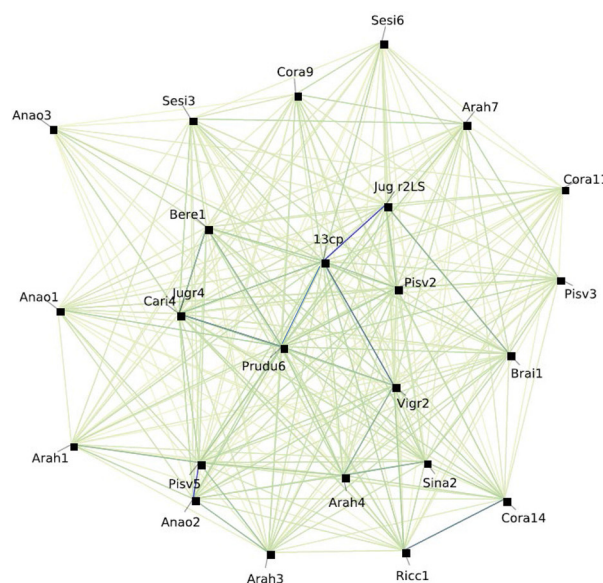


Fig. 2. A web of allergens with sequences similar to the 13cp shows this consensus represents a common feature among nut and seed allergens, many of which have overall low identity to one another. A. The table shows the sequences identified by PD search starting with the 13cp (CQRQEQQGRQQQ) listed from lowest PD value (most similar) to highest. The number of repeats indicate the number of sequences similar to the 13cp that are found within that allergen. B. The PD-graph was generated using the sequence in each allergen that was most similar (lowest PD) to the 13cp. Visible lines have inter-peptide PD < 10, where darker lines indicate higher pairwise similarity (PD values between 0 and 7).

2.7. Indirect ELISA

Peanut, walnut, almond, cashew, pistachio extracts and purified Jug r 1 (as negative control), in 0.1 M NaHCO₃ (pH 9.5), were added to a 96 well plate, at the indicated amounts and allowed to bind overnight at 4 °C. The plate was washed three times with PBST, blocked with 3 % dry milk in PBS for 2 h at 37 °C, and then washed with PBST three times. The extract-bound plates were incubated with cpAB (1:100) for one hour at 37 °C and then washed with PBST. The plates were then incubated for 1 h at 37 °C with a 1:5000 dilution in PBST of anti-rabbit IgG-HRP and then washed with PBST. The plates were incubated with the HRP substrate 3, 3', 5, 5'-tetramethylbenzidine (KPL, Gaithersburg, MD). The absorbance at 450 nm was determined in a Tecan Sunrise plate reader (Tecan US, San Jose, CA).

2.8. Competitive inhibition ELISA (ciELISA)

Walnut and almond extracts (at the amounts indicated) in 0.1 M NaHCO₃ (pH 9.5) were added to a 96 well plate and allowed to bind overnight at 4 °C. The plate was washed three times with PBST, blocked with 3 % dry milk in PBS for 2 h at 37 °C, and then washed with PBST three times. Increasing amounts of the cp31 or J2LS were mixed with equal volumes peanut and tree nut allergic patients' sera (1:5). After the mixtures were rotated at 4 °C for 2.5 h, they were added to the nut extract-bound plates and incubated for 1 h at 37 °C. The plates were washed with PBST and incubated with a 1:1500 dilution of HRP-conjugated mouse anti-human antibody (Sigma Chemical Company, St. Louis, MO) for one hour at 37 °C and bound HRP determined as described for indirect ELISA. *T*-tests were used to characterize how J2LS inhibition (Fig. 5C) and 31cp inhibition (Fig. 5D) influenced O.D. 450 intensity (a proxy for IgE binding) as compared to uninhibited controls in the competitive inhibition ELISA (ciELISA). Welch's unequal variances *t*-tests were performed in R (version 3.6.0; R Development Team,

available at www.r-project.org) using the *t*-test function.

2.9. Liquid chromatography–mass spectrometry (LC–MS/MS)

In-gel digestion with trypsin was carried out according to Chassaigne et al. (2009) with slight modifications. Bands were destained in 200 mM ammonium bicarbonate, washed with 50:50 acetonitrile:DH₂O, reduced with 10 mM DTT in 50 mM ammonium bicarbonate (pH 8.6) buffer for 30 min and alkylated in 55 mM iodoacetamide in 100 mM ammonium bicarbonate for 20 min. Trypsin (Promega, Sunnyvale, CA) in ammonium bicarbonate at 20 ng/μl was added to cover gel pieces and incubated overnight at 37 °C. The digests were analyzed via LC/MS/MS, using an Agilent 1200 LC system, an Agilent Chip-cube interface and an Agilent 6520 Q-TOF tandem mass spectrometer (Agilent Technologies, Santa Clara, CA, USA). Data files were transferred to an Agilent workstation equipped with Spectrum Mill software (Agilent Technologies). The raw MS/MS data files were extracted, sequenced, and searched against the NCBI nr, TrEMBL, and SwissProt databases.

3. Results

3.1. Determining a consensus sequence based on similar immunodominant IgE epitopes of walnut and peanut allergens

A sequence similar (PD = 5.9) to an immunodominant IgE epitope of Ara h 2 (DRRCQSQLER), QRQCQQRCE in the J2LS was identified by the peptide similarity tool of SDAP. These IgE epitopes were empirically confirmed to be cross-reactive between walnut and peanut allergic individuals (Maleki et al., 2011). Upon further examination, we found that J2LS consists of three long cysteine rich repeats (Fig. 1A) that can be aligned with IgE epitope rich regions of the PN-2S albumin allergen Ara h 2, as well as its related allergens, Ara h 6 (de Jong et al.,

2018) and Ara h 7 (Fig. 1B).

The PCPcon program was used to define a PCP-consensus region based on the three repeats in the J2LS (P1 (highlighted in green), P2 (highlighted in red) and P3 (highlighted in blue) in Fig. 1). The middle and most conserved region of this consensus, a 13 amino acid peptide (13cp, CQRQEQQGRRQQQ) was synthesized and used for antibody production. A peptide similarity (PD) search starting with the 13cp identified similar peptide sequences in many allergens from nuts and seeds (Fig. 2A). The predicted peptides span several protein families, including the 2S albumins: Ric I 1 (castor bean), Ber e 1 (Brazil nut), Cor a 14 (hazelnut), Ana o 3 (cashew); the 7S globulins: Ses i 3 (sesame), Ara h 1 (peanut), and the 11S globulins: Pru du 6 (almond) and Ana o 2 (cashew). Fig. 2A also shows that many allergens contain several discrete sequences with a low PD to the 13cp. In the most extreme example, Pru du 6 contains more than 10 non-overlapping sequences with PD values 4.95–9.25, spanning residues 96–339 (Fig. 2A). A PD-graph displays the interrelatedness of the peptides in each allergen with the lowest PD (Fig. 2B). As anticipated, the 13cp lies in the middle of the minimized distance plot, which links allergens from quite different PFAM, such as the vicilins (Ara h 1, Ara h 3, Pis v 3, etc.) and 2S albumins (Ara h7, Ric c 1, Cor a 14, etc.) The figure does not include the many similar sequences in the glutamine (Q) rich wheat gliadins and soy glycinins, which are also allergens.

3.2. IgE binds to sequences similar to the 13cp

We next determined how well patient IgE bound to sequences similar to the 13cp in different PN and TN allergens. To best determine their binding ability in the context of the original allergen, overlapping 15mer peptides offset by 5 aa were printed on microarray slides. The microarray was designed to identify IgE epitopes of PN and TN allergic individuals, so many peptides contain all or part of the similar areas. As

Fig. 3A shows, while the 13cp itself was only recognized by one serum, the naturally occurring peptides most similar to the 13cp in a number of PN and TN allergens are among those with the most frequent binding to IgE from PN, TN and PN + TN allergic sera. A PD-graph between the most IgE-reactive peptides (Fig. 3B) shows that once more, the 13cp lies at the center of the relationship web. This illustrates the interrelationship between these naturally occurring, IgE reactive sequences and the 13cp.

3.3. An antibody generated to the 13cp binds to allergens in nut extracts

To further test the consensus approach, a rabbit antibody against the 13cp (cpAB) was used to probe extracts of various nuts in western blots (Fig. 4). We hypothesized, based on validation of the PCP consensus method, that even though this polyclonal cpAB was made against a single peptide, it would bind proteins containing regions of similar PCPs (i.e. sequences with low PD to the 13cp). Despite the high amounts of protein loaded in the gel lanes, the cpAB showed both strong and weak reactivity with specific protein bands. These were excised and the proteins identified by LC-MS/MS, except for bands we had previously identified in similar experiments, such as those for the 2S albumins of peanut. The identified proteins (Fig. 4C shows the % coverage of the tryptic fragments) correlated well with allergens predicted to contain multiple copies of sequences with low PD to the 13cp. Binding of the cpAB to the nut extracts was quantified in an indirect ELISA (Fig. 4D). The binding order (almond > pistachio > cashew/walnut/peanut) was consistent with allergens with the most repeats similar to the 13cp (Fig. 2A), particularly Pru du 6 (prunin) from almonds, Pis v 2 from pistachio and the J2LS.

A. IgE binding to peptides on the microarray that contain aa sequences similar to the 13cp.

No. on D-Graph	Source allergen	Peptide Sequence present on the microarray	# of patient IgE binding on microarray		
			PN	TN	PN+TN
1	13cp	CQRQEQQGRRQQQ	0	1	0
2	Ana o 1	ELKQCKHQCKVQRQY	3	1	1
3	Ana o 3	RCRCQNLEQMVRQLQ	4	0	2
4	Cor a 14	SCREQAQRQQNLNQC	2	0	0
5	Ara h 2	ELQGDRCQSQLEIA	5	5	10
6	Ara h 2b	RLQGRQEQEQFKREL	1	1	1
7	Ara h 1	AGGEQEERQRRWST	6	1	3
8	Ara h 1b	ELVAVRKEQQQRGR	4	2	0
9	Cor a 14b	SCREQAQRQQNLNQC	1	1	0
10	Ara h 3	IFPGCPRHYYEPTQ	7	3	2
11	Ara h 7	MRRRVEQEQQEQDE	2	0	1
12	Ara h 11	MAEALYGGRRQREQ	6	1	4
13	Pis v 1	VDKKCRQCQNLEQMVK	6	2	0
14	Pis v 3	LKQCKHQCKVQRQYD	3	2	0
15	Jug r 2	QRRGQEQLCRRRCE	3	4	2
16	Jug r 2b	CQEYCRRGQGGRQQ	4	2	2
17	Jug r 2c	PEASPRRESRGREE	2	1	0
18	Jug r 4	QGQEQEYQHRRQQQR	2	0	1
19	Jug r 4b	RQQRRQRPGEHGQQ	2	0	0

B.

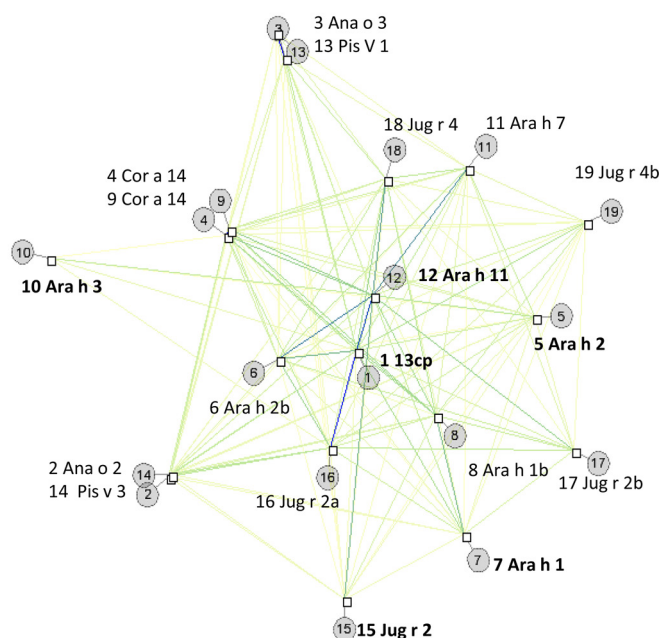


Fig. 3. IgE from sera of patients with peanut (PN) and tree nut (TN) allergies recognize peptides from PN and TN allergens similar to the 13cp. A. Fifteen amino acid peptides that were recognized by IgE in sera from at least one PN or TN patient. The last three columns indicate the number of patients allergic to PN, TN PN + TN whose sera showed significant IgE binding to peptides containing the identified sequences from Fig. 2A, according to microarray analysis. The 13cp and peptides recognized by at least 9 of the nut allergic patient sera are in bold. Sequences that are nearly identical in different allergens, such as those of cashew (Ana o 1 and Ana o 3) and pistachio (Pis v 1 and Pis v 3), which occur in different areas of the microarray and in different context in the individual peptides, have similar reactivity. B. The PD-graph shows the relationship between the 13cp and the reactive peptides in panel A. Individual lines represent inter-peptide PD values between 2 and 10 where the thicker the line, the higher the similarity (lower the PD value). Dark lines are PD values up to 6.

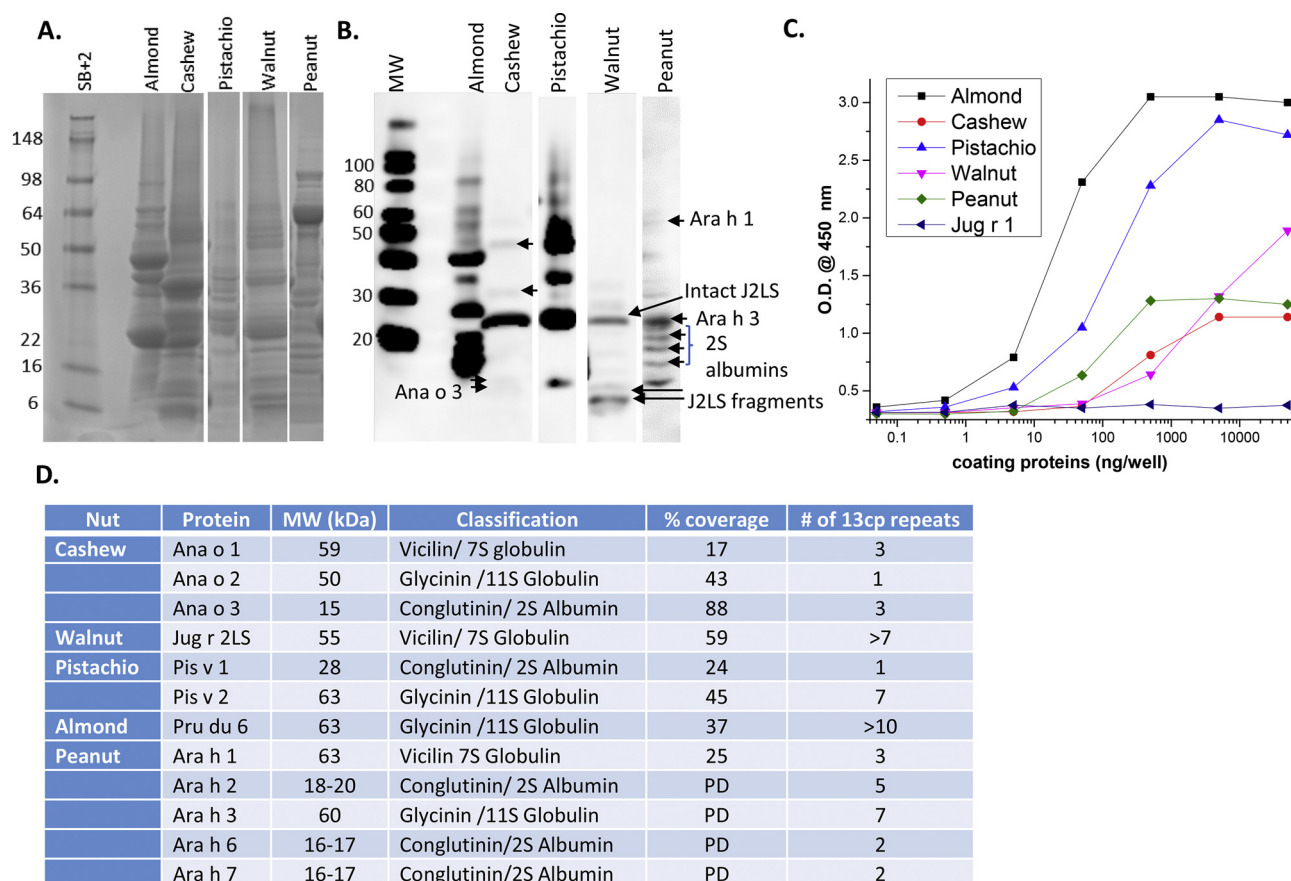


Fig. 4. The cpAB detects allergens with repeats similar to the 13cp in peanut and tree nut extracts. A. Coomassie-stained SDS-PAGE of the indicated nut extract proteins separated on a 4–20 % SDS-PAGE gel. B. The consensus peptide antibody (cpAB) binds to distinct bands at various molecular weights (MW) as shown by the western blot. C. ELISA measuring direct binding of cpAB to nut extracts where the binding is measured by absorbance at 450 nm (y-axis) as a function of the increasing concentration of proteins from the indicated nut extract coating the wells of the ELISA plate (x-axis). The purified Jug r 1 was used as a negative control. D. Table showing proteins excised, subjected to tryptic digest and identified by LC-MS/MS, as well as the % coverage of the tryptic fragments found. The number of 13cp repeats pertains to the number of sequences within the allergen that have PD score less than 9 to the 13cp in a SDAP peptide similarity search. PI indicates these bands were previously identified.

3.4. cpAB mimics the binding pattern of allergic patient IgE to almond and walnut extracts

The cpAB binding pattern on the western blots mimicked that we observed for the binding of IgE from almond allergic sera to almond extracts (Fig. 5). This, coupled with our other assays, strongly suggests that at least some fraction of IgE in the patient sera is specific for the 13cp repeats. This was further confirmed by ciELISA, which showed recombinant J2LS, which contains three long repeat areas similar to the consensus (Fig. 1) interfered with IgE binding to nut extracts, using sera from three almond and walnut allergic individuals (Fig. 5C). The J2LS inhibited IgE binding significantly for all three sera according to the calculated p-values (Fig. 5C). The 31cp significantly inhibited the binding of a walnut allergic patient sera (patient #3) to walnut extract with the calculated p-values indicated in Fig. 5D. The inhibition observed for almond extracts, at the concentration of the 31mer achievable in this experiment, was either marginally or not significant.

4. Discussion

The prevailing theory is that cross-reactive allergens resemble one another closely in sequence and/or structure (Aalberse et al., 2001; Aalberse and Stadler, 2006; Soman et al., 2000). However, experimental data has shown that there is little relationship between overall % identity of known food allergens and their ability to bind IgE from patient sera or cause clinical cross reactivity (Schein et al., 2007;

Smeekens et al., 2018). This study shows the power of combining computational methods with IgE binding assays to determine common characteristics of distantly related allergens. Our results show that a PCP-consensus peptide, calculated initially based on repeats in the J2LS, which has high similarity to an IgE epitope rich region of peanut 2S albumins, can be used to define similar sequences in many seemingly unrelated allergens (Figs. 1 & 2). By microarray (Fig. 3) and ciELISA assays (Fig. 4), using sera from patients severely allergic to PN and TN, we experimentally confirm that peptides with low PD to the 13cp bind to serum IgE. Antibodies to 13cp also recognize the predicted allergens in nut extracts. Indeed, at least for almond extract, the pattern of protein binding by the cpAB in a western blot is essentially identical to that of patient IgE binding. This strongly suggests that these similar sequences play a role in clinically observed cross-reactivity (Schein et al., 2007).

Our finding, that the cpAB recognized whole allergens that contain repeats similar to the consensus in nut extracts (Figure 2A & 4) indicates that our PCP-based computational methods can indeed detect common regions in allergens from many sources. Further, our data indicate that we are able to identify sequences that react with patient IgE consistent with their similarity to the consensus. We have thus identified common, cross-reactive regions in these allergens that could account for both the high reactivity of the allergens in which they occur, and for the clinically observed cross-sensitivity between PN, TN and other plant foods. Other common areas of allergens, such as the PCP-motifs we defined previously (Lu et al., 2018) may also contribute to

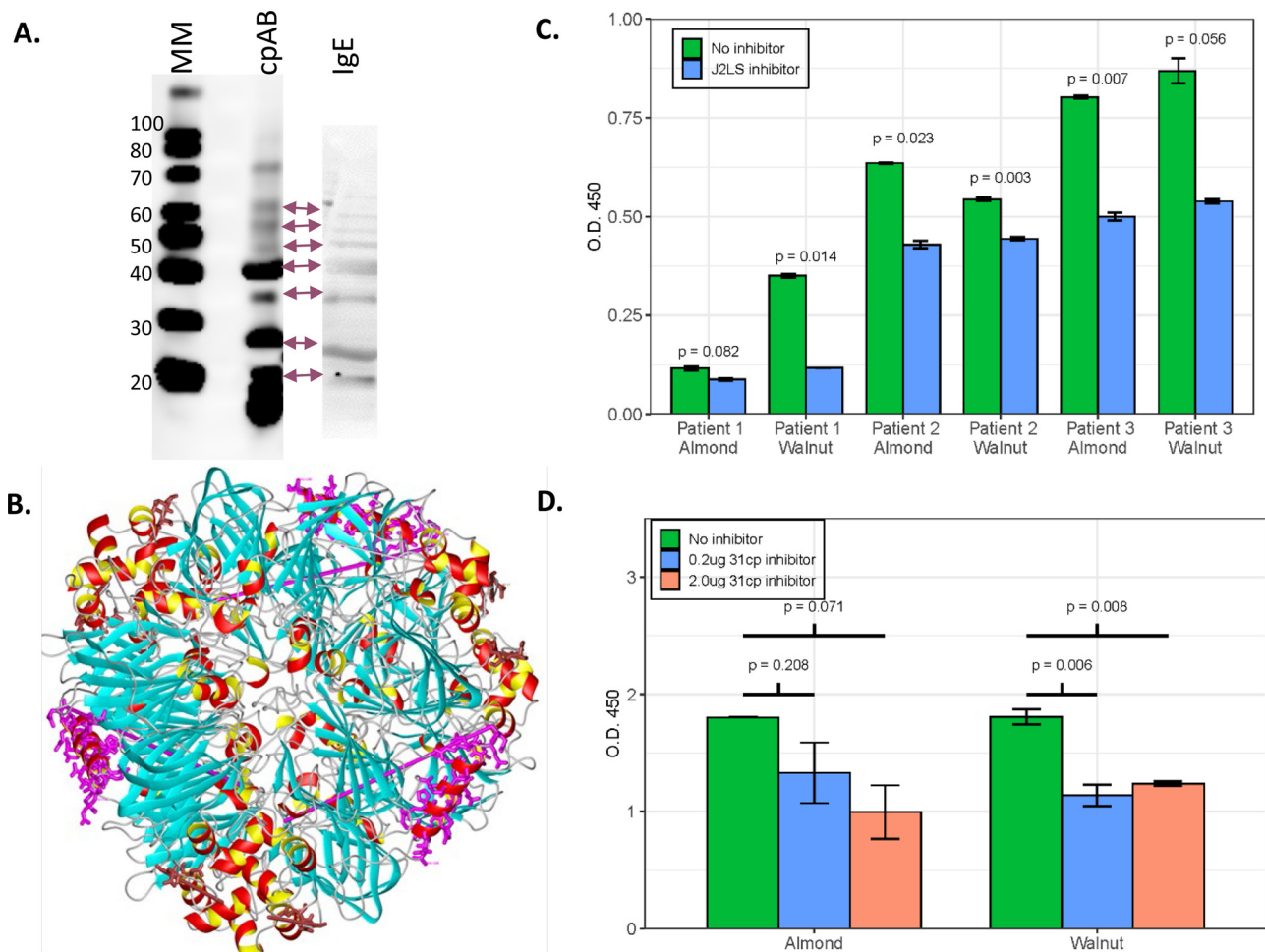


Fig. 5. The binding of cpAB mimics the binding of IgE from sera of almond allergic patients to almond extract. A. Comparison of the protein bands recognized by the cpAB in a western blot with an extract of almond to that bound by IgE in sera from a representative almond allergic patient (lane on right). B. The hexamer of the crystal structure of Pru du 6, where the magenta areas mark the locations of the 13cp repeats (straight lines indicate strings of missing residues in these areas of unresolved structure). The 13cp repeats mark surface areas of high mobility, whose atomic structure could not be ascertained from the electron density map. C. ciELISA in which J2LS competes for IgE binding by 3 different patient sera to almond or walnut extracts. Shown are the average O.D. 450 of IgE binding with standard errors (y-axis). Results of Welch's *t*-tests used to test for differences in mean O.D. 450 between no inhibitor (green bars) and J2LS inhibited samples (blue bars) are shown. D. ciELISA in which indicated amounts of 31cp compete for IgE binding by serum of patient #3 with almond and walnut extracts. This assay was performed for one patient due to limited amounts of synthetic 31cp. Results of Welch's *t*-tests used to test for differences in mean O.D. 450 between no inhibitor (green bar) and 0.2 ug 31cp (blue bar) or 2.0 ug 31cp (red bar) are shown.

cross-reactivity and the severity of responses in sensitive individuals.

The occurrence of repeat sequences, similar to the 13cp, in allergens from plants that evolved from one another more than 300 million years ago (Fisher and Nawrocki, 2016) suggests that they are ancient in origin, and may serve a functional role in the seed. For example, proteins containing cysteine rich repeats, isolated from macadamia nuts, were shown to have antimicrobial properties (Marcus et al., 1997, 1999). Even more repeats similar to those we identified may be present in the edible plants that cause allergic reactions, due to transposons and their polyploidy (Nicolas et al., 2016). Their presence in many different allergenic proteins within the same food (i.e. in five different PN allergens, Ara h 1, 2, 3 6 and 7) can also account for why peanuts are so reactive, and why removing a single protein may not significantly reduce allergenicity (Chen et al., 2011; Hazebrouck et al., 2019).

Whether these repeats play a role in inducing allergic responses may be dependent on their degree of exposure in the plant protein. For example, a crystal structure of Pru du 6 indicates the areas where the 13cp repeats occur should be surface exposed (Fig. 5B). They are probably also very flexible in structure, as these regions were not clearly resolved in the electron density maps. This suggests the repeats are located in areas of high mobility that are more likely to be susceptible to digestive

enzymes (Maleki et al., 2014), which, in turn, will affect their ability to interact with the immune system. The sequence of the consensus itself, which is both glutamine rich and positively charged, may also indicate why the related peptides are detected by an activated immune system. Glutamine repeats (polyQ) are known to contribute to transcription factor activity, but are also implicated in mammalian disease (Schein, 2019). For example, polyQ proteins in wheat contribute to celiac disease (gluten sensitivity) and allergenicity and contain many repeats similar to our consensus. The secondary and tertiary structure surrounding the peptides is likely to play a role in their reactivity. The 13mer and the reactive peptides discussed in this paper are too short to form the disulfides anticipated from the NMR structure of Ara h 6 (PDB file 1W2Q) (Lehmann et al., 2006) and the crystal structure of Ara h 2 (PDB file 3OB4) (Mueller et al., 2011). However, we anticipate that by careful preparation of the input alignment, the properties of residues that contribute to correct disulfide bonding and other features essential to a stable 3D-structure will be maintained.

These results are important for both diagnosis and therapy of severe allergy. Recombinant J2LS, which contains three long sequences similar to our 31cp, inhibited the binding of IgE from three nut allergic sera to total almond and walnut extracts (Fig. 5C). This, coupled with the fact

that the 13cp is not as reactive with patient IgE as the natural sequences similar to it, suggests that further decomposition of the component repeats of the J2LS should be done. This could provide a novel way to inhibit IgE binding, and aid in designing treatments or for assessing the specificity of patient responses. Analogous to ongoing trials with specifically selected regions of grass pollen allergens (Niederberger et al., 2018), peptides from this region could aid in safe desensitization without the risk of inducing anaphylaxis.

5. Conclusions

Overall, these results indicate how a combined experimental and bioinformatics approach to protein and peptide recognition by the immune system, can be useful for identifying new epitopes and motifs thereof. Our previous studies (Ivanciuc et al., 2009b; Maleki et al., 2011; Schein et al., 2005a) validated the peptide similarity search based on PD values for use in predicting areas of homologous proteins that could be cross-reactive. The ramifications for this work are significant for the detection of potentially cross-reactive allergenic proteins, and for generating or selecting new diagnostic and immunotherapeutic peptides or proteins.

Author statements

All authors have approved the submission of this manuscript. The results have not been previously published, except as abstracts and poster presentations, and are not being considered for publication in another journal.

Source of funding

This work was supported by EPA Star award RE-83406601-0 to CHS; NIH1 R01AI064913-01 to Werner Braun and CHS; EPA RD834823-01 to Werner Braun/CHS; USDA – ARS 58-6435-9-405 to CHS; and funding to SJM & BKH from the USDA-ARS6054-43440-046-00D; USDA-NIFA and National Peanut Board (GRANT12229460).

Author contributions

JBN Performed majority of the experiments and wrote a significant part of the manuscript; BKH, HC designed and performed the microarray experiments; BAB wrote the D-graph program and prepared figures; SAYG performed statistical analysis; CHS co-directed the study, supervised the computational work, helped prepare figures and the manuscript; SJM co-directed the study design and implementation and, helped prepare figures and manuscript; SJM, BKH and CHS funded the project.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

We are grateful to the developers and maintainers of the PD scale and related tools within the SDAP web server: Ovidiu Ivanciuc PhD; Werner Braun, PhD (UTMB, Galveston); and helpful discussions with Suzanne S. Teuber, MD, UC Davis.

References

Aalberse, R.C., Stadler, B.M., 2006. In silico predictability of allergenicity: from amino acid sequence via 3-D structure to allergenicity. *Mol. Nutr. Food Res.* 50, 625–627.
Aalberse, R.C., Akkerdaas, J., van Ree, R., 2001. Cross-reactivity of IgE antibodies to allergens. *Allergy* 56, 478–490.
Bowen, D.M., Lewis, J.A., Lu, W., Schein, C.H., 2012. Simplifying complex sequence

information: a PCP-consensus protein binds antibodies against all four Dengue serotypes. *Vaccine* 30, 6081–6087.
Bublin, M., Breiteneder, H., 2014. Cross-reactivity of peanut allergens. *Curr. Allergy Asthma Rep.* 14, 426.
Chassaing, H., Tregat, V., Norgaard, J.V., Maleki, S.J., van Hengel, A.J., 2009. Resolution and identification of major peanut allergens using a combination of fluorescence two-dimensional differential gel electrophoresis, Western blotting and Q-TOF mass spectrometry. *J. Proteomics* 72, 511–526.
Chen, X., Zhuang, Y., Wang, Q., Moutsoglou, D., Ruiz, G., Yen, S.E., Dreskin, S.C., 2011. Analysis of the effector activity of Ara h 2 and Ara h 6 by selective depletion from a crude peanut extract. *J. Immunol. Methods* 372, 65–70.
Chen, J., Li, R., Xia, Y., Bai, G., Guo, P., Wang, Z., Zhang, H., Siddique, K.H.M., 2017. Development of EST-SSR markers in flowering Chinese cabbage (*Brassica campestris* L. ssp. *chinensis* var. *utilis* Tsen et Lee) based on de novo transcriptomic assemblies. *PLoS One* 12, e0184736.
Danecek, P., Schein, C.H., 2010. Flavitrack analysis of the structure and function of West Nile non-structural proteins. *Int. J. Bioinform. Res. Appl.* 6, 134–146.
Danecek, P., Lu, W., Schein, C.H., 2010. PCP consensus sequences of flaviviruses: correlating variance with vector competence and disease phenotype. *J. Mol. Biol.* 396, 550–563.
de Jong, G.A.H., Jayasena, S., Johnson, P., Marsh, J., Apostolovic, D., van Hage, M., Nordlee, J., Baumert, J., Taylor, S.L., Roucairol, C., de Jongh, H., Koppelman, S.J., 2018. Purification and characterization of naturally occurring post-translationally cleaved Ara h 6, an allergen that contributes substantially to the allergenic potency of peanut. *J. Agric. Food. Chem.* 66, 10855–10863.
de Leon, M.P., Glaspole, I.N., Drew, A.C., Rolland, J.M., O'Hehir, R.E., Suphioglu, C., 2003. Immunological analysis of allergenic cross-reactivity between peanut and tree nuts. *Clin. Exp. Allergy* 33, 1273–1280.
de Leon, M.P., Drew, A.C., Glaspole, I.N., Suphioglu, C., O'Hehir, R.E., Rolland, J.M., 2007. IgE cross-reactivity between the major peanut allergen Ara h 2 and tree nut allergens. *Mol. Immunol.* 44, 463–471.
Fisher, A.E., Nawrocki, A.M., 2016. An evolutionary perspective on human cross-sensitivity to tree nut and seed allergens. *ALISO: J. Syst. Evol. Bot.* 33, 91–110.
Glaspole, I.N., de Leon, M.P., Prickett, S.R., O'Hehir, R.E., Rolland, J.M., 2011. Clinical allergy to hazelnut and peanut: identification of T cell cross-reactive allergens. *Int. Arch. Allergy Immunol.* 155, 345–354.
Hazebrouck, S., Guillon, B., Paty, E., Dreskin, S.C., Adel-Patient, K., Bernard, H., 2019. Variable IgE cross-reactivity between peanut 2S-albumins: the case for measuring IgE to both Ara h 2 and Ara h 6. *Clin. Exp. Allergy* 49, 1107–1115.
Ivanciuc, O., Schein, C.H., Braun, W., 2002. Data mining of sequences and 3D structures of allergenic proteins. *Bioinformatics* 18, 1358–1364.
Ivanciuc, O., Mathura, V., Midoro-Horiuti, T., Braun, W., Goldblum, R.M., Schein, C.H., 2003a. Detecting potential IgE-reactive sites on food proteins using a sequence and structure database, SDAP-food. *J. Agric. Food. Chem.* 51, 4830–4837.
Ivanciuc, O., Schein, C.H., Braun, W., 2003b. SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res.* 31, 359–362.
Ivanciuc, O., Oezguen, N., Mathura, V.S., Schein, C.H., Xu, Y., Braun, W., 2004. Using property based sequence motifs and 3D modeling to determine structure and functional regions of proteins. *Curr. Med. Chem.* 11, 583–593.
Ivanciuc, O., Garcia, T., Torres, M., Schein, C.H., Braun, W., 2009a. Characteristic motifs for families of allergenic proteins. *Mol. Immunol.* 46, 559–568.
Ivanciuc, O., Midoro-Horiuti, T., Schein, C.H., Xie, L., Hillman, G.R., Goldblum, R.M., Braun, W., 2009b. The property distance index PD predicts peptides that cross-react with IgE antibodies. *Mol. Immunol.* 46, 873–883.
Kaur, S., Panesar, P.S., Bera, M.B., Kaur, V., 2015. Simple sequence repeat markers in genetic divergence and marker-assisted selection of rice cultivars: a review. *Crit. Rev. Food Sci. Nutr.* 55, 41–49.
Kulis, M., Pons, L., Burks, A.W., 2009. In vivo and T cell cross-reactivity between walnut, cashew and peanut. *Int. Arch. Allergy Immunol.* 148, 109–117.
Lehmann, K., Schweimer, K., Reese, G., Randow, S., Suhr, M., Becker, W.M., Vieths, S., Rosch, P., 2006. Structure and stability of 2S albumin-type peanut allergens: implications for the severity of peanut allergic reactions. *Biochem. J.* 395, 463–472.
Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R.J., Ma, Z., Shang, H., Ma, X., Wu, J., Liang, X., Huang, G., Percy, R.G., Liu, K., Yang, W., Chen, W., Du, X., Shi, C., Yuan, Y., Ye, W., Liu, X., Zhang, X., Liu, W., Wei, H., Wei, S., Huang, G., Zhang, X., Zhu, S., Zhang, H., Sun, F., Wang, X., Liang, J., Wang, J., He, Q., Huang, L., Wang, J., Cui, J., Song, G., Wang, K., Xu, X., Yu, J.Z., Zhu, Y., Yu, S., 2015. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Bio-Technology* 33, 524.
Lu, W., Negi, S.S., Schein, C.H., Maleki, S.J., Hurlburt, B.K., Braun, W., 2018. Distinguishing allergens from non-allergenic homologues using Physical-Chemical Property (PCP) motifs. *Mol. Immunol.* 99, 1–8.
Luo, M.-C., You, F.M., Li, P., Wang, J.-R., Zhu, T., Dandekar, A.M., Leslie, C.A., Aradhya, M., McGuire, P.E., Dvorak, J., 2015. Synteny analysis in Rosids with a walnut physical map reveals slow genome evolution in long-lived woody perennials. *BMC Genomics* 16, 707.
Maleki, S.J., Teuber, S.S., Cheng, H., Chen, D., Comstock, S.S., Ruan, S., Schein, C.H., 2011. Computationally predicted IgE epitopes of walnut allergens contribute to cross-reactivity with peanuts. *Allergy* 66, 1522–1529.
Maleki, S.J., Teuber, S.S., Schein, C.H., 2012. Similar repeated sequences may account for cross-reactions caused by many different nuts. *J. Allergy Clin. Immunol.* 129, Ab233–Ab233.
Maleki, S.J., Schmitt, D.A., Galeano, M., Hurlburt, B.K., 2014. Comparison of the digestibility of the major peanut allergens in thermally processed peanuts and in pure form. *Food* 3, 290–303.
Maloney, J.M., Rudengren, M., Ahlstedt, S., Bock, S.A., Sampson, H.A., 2008. The use of

- serum-specific IgE measurements for the diagnosis of peanut, tree nut, and seed allergy. *J. Allergy Clin. Immunol.* 122, 145–151.
- Marcus, J.P., Goulter, K.C., Green, J.L., Harrison, S.J., Manners, J.M., 1997. Purification, characterisation and cDNA cloning of an antimicrobial peptide from *Macadamia integrifolia*. *Eur. J. Biochem.* 244, 743–749.
- Marcus, J.P., Green, J.L., Goulter, K.C., Manners, J.M., 1999. A family of antimicrobial peptides is produced by processing of a 7S globulin protein in *Macadamia integrifolia* kernels. *Plant J* 19, 699–710.
- Mueller, G.A., Gosavi, R.A., Pomes, A., Wunschmann, S., Moon, A.F., London, R.E., Pedersen, L.C., 2011. Ara h 2: crystal structure and IgE binding distinguish two subpopulations of peanut allergic patients by epitope diversity. *Allergy* 66, 878–885.
- Nicolas, J., Peterlongo, P., Tempel, S., 2016. Finding and characterizing repeats in plant genomes. *Methods Mol. Biol.* 1374, 293–337.
- Niederberger, V., Neubauer, A., Gevaert, P., Zidarn, M., Worm, M., Aberer, W., Malling, H.J., Pfaar, O., Klimek, L., Pflutzner, W., Ring, J., Darsow, U., Novak, N., Gerth van Wijk, R., Eckl-Dorna, J., Focke-Tejkl, M., Weber, M., Muller, H.H., Klinger, J., Stolz, F., Breit, N., Henning, R., Valenta, R., 2018. Safety and efficacy of immunotherapy with the recombinant B-cell epitope-based grass pollen vaccine BM32. *J. Allergy Clin. Immunol.* 142 497–509 e499.
- Oezguen, N., Zhou, B., Negi, S.S., Ivanciuc, O., Schein, C.H., Labesse, G., Braun, W., 2008. Comprehensive 3D-modeling of allergenic proteins and amino acid composition of potential conformational IgE epitopes. *Mol. Immunol.* 45, 3740–3747.
- Schein, C.H., 2019. Polyglutamine repeats in viruses. *Mol. Neurobiol.* 56, 3664–3675.
- Schein, C.H., Ivanciuc, O., Braun, W., 2005a. Common physical-chemical properties correlate with similar structure of the IgE epitopes of peanut allergens. *J. Agric. Food. Chem.* 53, 8752–8759.
- Schein, C.H., Zhou, B., Oezguen, N., Mathura, V.S., Braun, W., 2005b. Molego-based definition of the architecture and specificity of metal-binding sites. *Proteins* 58, 200–210.
- Schein, C.H., Ivanciuc, O., Braun, W., 2007. Bioinformatics approaches to classifying allergens and predicting cross-reactivity. *Immunol. Allergy Clin. North Am.* 27, 1–27.
- Schein, C.H., Ivanciuc, O., Midoro-Horiuti, T., Goldblum, R.M., Braun, W., 2010. An allergen portrait gallery: representative structures and an overview of IgE binding surfaces. *Bioinform. Biol. Insights* 4, 113–125.
- Schein, C.H., Bowen, D.M., Lewis, J.A., Choi, K., Paul, A., van Noort, G.J.V., Lu, W.Z., Filippov, D.V., 2012. Physicochemical property consensus sequences for functional analysis, design of multivalent antigens and targeted antivirals. *BMC Bioinf.* 13, S9.
- Schein, C.H., Ye, M., Paul, A.V., Oberste, M.S., Chapman, N., van der Heden van Noort, G.J., Filippov, D.V., Choi, K.H., 2015. Sequence specificity for uridylation of the viral peptide linked to the genome (VPg) of enteroviruses. *Virology* 484, 80–85.
- Sicherer, S.H., Munoz-Furlong, A., Sampson, H.A., 2003. Prevalence of peanut and tree nut allergy in the United States determined by means of a random digit telephone survey: a 5-year follow-up study. *J. Allergy Clin. Immunol.* 112, 1203–1207.
- Smeekens, J.M., Bagley, K., Kulis, M., 2018. Tree nut allergies: allergen homology, cross-reactivity, and implications for therapy. *Clin. Exp. Allergy* 48, 762–772.
- Soman, K.V., Midoro-Horiuti, T., Ferreon, J.C., Goldblum, R.M., Brooks, E.G., Kurosky, A., Braun, W., Schein, C.H., 2000. Homology modeling and characterization of IgE binding epitopes of mountain cedar allergen Jun a 3. *Biophys. J.* 79, 1601–1609.
- Uotila, R., Kukkonen, A.K., Blom, W.M., Remington, B., Westerhout, J., Pelkonen, A.S., Makela, M.J., 2018. Component-resolved diagnostics demonstrates that most peanut-allergic individuals could potentially introduce tree nuts to their diet. *Clin. Exp. Allergy* 48, 712–721.
- Wang, H., Khera, P., Huang, B., Yuan, M., Katam, R., Zhuang, W., Harris-Shultz, K., Moore, K.M., Culbreath, A.K., Zhang, X., Varshney, R.K., Xie, L., Guo, B., 2016. Analysis of genetic diversity and population structure of peanut cultivars and breeding lines from China, India and the US using simple sequence repeat markers. *J. Integr. Plant Biol.* 58, 452–465.

Glossary

- PD: P
SDAP: S
PCP: P
cp: C
13cp/31cp: 13 aa or 31 aa consensus peptide
cpAB: C
IgE: Immunoglobulin E
sIgE: S
LC-MSMS: Liquid Chromatography-Mass Spectrometry
PFAM: Protein family
PN: Peanut PN
TN: Tree nut
ELISAE: enzyme-linked immune assay
ciELISA: C
J2LS: Jug r 2 leader sequence