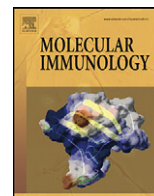




Contents lists available at ScienceDirect

Molecular Immunology

journal homepage: www.elsevier.com/locate/molimm



Linear IgE-epitope mapping and comparative structural homology modeling of hazelnut and English walnut 11S globulins

Jason M. Robotham^a, Gregg G. Hoffman^a, Suzanne S. Teuber^b, Kirsten Beyer^c, Hugh A. Sampson^d, Shridhar K. Sathe^e, Kenneth H. Roux^{a,*}

^a Department of Biological Science and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL, USA

^b Department of Internal Medicine, School of Medicine, University of California, Davis, CA, USA

^c Department of Pediatric Pneumology and Immunology, University Children's Hospital Charité, Berlin, Germany

^d Department of Pediatric Allergy, Mount Sinai School of Medicine, New York, NY, USA

^e Department of Nutrition, Food and Exercise Sciences, Florida State University, Tallahassee, FL, USA

ARTICLE INFO

Article history:

Received 5 May 2009

Accepted 16 June 2009

Available online xxx

Keywords:

Jug r 4

Cor a 9

Food allergy

Walnut allergen

Hazelnut allergen

Tree nut allergy

11S globulin

Legumin

Epitope mapping

Homology modeling

ABSTRACT

Allergic reactions to walnuts and hazelnuts can be serious. The 11S globulins (legumins) have been identified as important allergens in these and other nuts and seeds. Here we identify the linear IgE-binding epitopes of walnut and hazelnut 11S globulins, and generate 3D 11S globulin models to map the locations of the epitopes for comparison to other allergenic homologues. Linear IgE-epitope mapping was performed by solid-phase overlapping 15-amino acid peptides probed with IgE from pooled allergic human sera. Several walnut (Jug r 4) and hazelnut (Cor a 9) 11S globulin peptides with reactivity to patient IgE were identified. Comparative alignment with cashew (Ana o 2), peanut (Ara h 3), and soybean G1 (Gly m 6.0101) and G2 (Gly m 6.0201) allergenic homologues revealed several shared allergenic 'hot spots'. Homology modeling was performed based on the atomic structure of the soybean glycinin. Surface map comparisons between the tree nut and peanut homologues revealed structural motifs that could be important for IgE elicitation and binding and show that, contrary to predictions, the reactive epitopes are widely distributed throughout the monomeric subunits, both internally and externally, including regions occluded by quaternary subunit association. These findings reveal structural features that may be important to allergenicity and cross-reactivity of this protein class.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Allergies to tree nuts and peanuts can be severe, are generally not outgrown, and appear to be increasing in prevalence (Bock et al., 2001; Sicherer et al., 2003). Sensitivity to walnuts, cashews and hazelnut are the first, second, and sixth most frequent tree nut allergies in the US, respectively (Sicherer et al., 2003). Most identified nut allergens are seed storage proteins, many of which are members of the evolutionarily related cupin family 11S and 7S globulins, also termed legumins and vicilins, respectively (Breiteneder and Radauer, 2004). The 11S globulins are synthesized as single polypeptides which are most often post-translationally cleaved into acidic (MW ~ 50–60 kDa) and basic (MW ~ 20 kDa) subunits that remain joined by an intermolecular disulfide bond (Shewry et al.,

2003). These proteins are typically deposited in high quantities in the maturing nut as non-glycosylated trimers or dimers of trimers (hexamers) with molecular masses of up to 450 kDa (Shewry et al., 2003).

Recombinant 11S proteins of walnut (Jug r 4) (Wallowitz et al., 2006), hazelnut (Cor a 9) (Beyer et al., 2002), cashew (Ana o 2) (Wang et al., 2003), and peanut (Ara h 3) (Rabjohn et al., 1999) have been identified as important allergens. The amino acid (aa) sequence of this family of proteins is fairly well conserved, and IgE cross-reactivity between tree nuts and, to a lesser extent, between tree nuts and peanuts has been reported (Poltronieri et al., 2002; Asero et al., 2004; Wallowitz et al., 2006; de Leon et al., 2003; Goetz et al., 2005; de Leon et al., 2005). Our recent studies show that the 11S globulins of these nuts are likely to be involved in this cross-reactivity (Wallowitz et al., 2006, 2007). Based on a comparison of peanut Ara h 3 (Rabjohn et al., 1999) and cashew Ana o 3 (Wang et al., 2003) linear epitopes, aa homologies and molecular modeling, Barre et al. (2007) predicted that certain peptide segments in the walnut and hazelnut would be IgE targets. Here, we identify the linear IgE-binding legumin epitopes of hazelnut Cor a 9 and English walnut Jug r 4, compare them with those previously identi-

Abbreviations: MW, molecular weight; aa, amino acid.

* Corresponding author at: Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295, USA. Tel.: +1 850 644 5037; fax: +1 850 645 8447.

E-mail address: roux@bio.fsu.edu (K.H. Roux).

fied on cashew Ana o 2 and peanut Ara h 3, and, through homology modeling, generate 3D epitope maps of all 4 allergens for structural comparisons.

2. Methods

2.1. Human sera

Blood samples were drawn after informed consent from patients with self-reported reactions to walnut and hazelnut and the sera stored at -70°C until use. The study was approved by the relevant institutional review boards. The presence of walnut- and hazelnut-reactive IgE was confirmed by CAP FEIA assay (Phadia, Inc., Uppsala, Sweden) and specific Jug r 4 or Cor a 9 reactivity was confirmed by immunoblotting against nut protein extract. Clinical characteristics of the subjects are shown in Table 1.

2.2. Solid-phase peptide (SPOTs) synthesis and binding to IgE

SPOTs membranes displaying either the 508 aa sequence of Jug r 4 or the 515 aa sequence of Cor a 9, in the form of 62 or 63 15-aa peptides, respectively, each offset by 8 residues, were synthesized on derivatized cellulose sheets and probed as previously described (Wang et al., 2003).

2.3. Sequence analysis and molecular modeling

Primary sequences were aligned using BLAST 2.0. For modeling, aa sequences were aligned with the sequence of the soybean glycinin A3B4 homohexamer (Adachi et al., 2003) (PDB 1od5, <http://www.rcsb.org/pdb/home/home.do>) using CLUSTALX (Thompson et al., 1997) and adjusted manually to allow modeling of target sequence regions not resolved in the crystal structure. Models were constructed by submitting the threaded and aligned PDB files to the automatic homology server SWISS-MODEL (Schwede

et al., 2003). The resultant PDB models were subjected to successive rounds of Steepest Descent energy minimization (until ΔE was <0.050 kJ/M) using GROMOS96 (van Gunsteren and Berendsen, 1987) to relieve local conflicts and bond length anomalies. A single monomer from each target was selected and superimposed over the remaining two subunits to impose symmetry in the final trimer model using THESEUS (Theobald and Wuttke, 2006; Theobald and Wuttke, 2008). Final models for imaging were created using PyMol (<http://www.pymol.org>) by merging a solid surface model of the templated regions with a ribbon diagram of the un-templated regions. Swiss-PdbViewer v4.0.1 (<http://www.expasy.org/spdbv/>) was used to generate electrostatic charge models. Solvent-exposed residues (1.4 Å probe) on the monomer, trimer and hexamer were determined based on modeling of the soybean proglycinin A3B4 homohexamer (PDB 1od5) and analysis using the GETAREA program (<http://curie.utmb.edu/getarea.html>).

3. Results

3.1. Jug r 4 linear epitopes

The aa sequence of recombinant (r)Jug r 4 was screened for IgE-binding linear epitopes by probing 62 overlapping solid-phase synthetic peptides with sera from 17 walnut-allergic (Table 2), rJug r 4-IgE reactive patients (Wallowitz et al., 2006). Each was assigned to one of four pools (3–4 sera each). As presented in Table 2, the 17 serum samples, distributed among four pools, collectively reacted strongly with 3, moderately with 7, and weakly with 12 peptides. It should be noted that differences in the intensity of the reaction may reflect the affinity of the reactants, the relative amount of IgE antibody present in the sera or a combination of the two parameters. Many of the reactive peptides partially overlap adjacent reactive peptides and thus may together represent single epitopes.

Table 1
Characteristics of walnut and hazelnut-allergic subjects.

No. ^a	Sex	Age	Allergy to walnut (W)/hazelnut (H)	IgE reactivity on immunoblot to W or H	Other atopy	Other known food reactivity
1	M	25	W, H	W = weak	asthma	cashew, pecan
3	F	26	W	W = weak	AD, AR asthma	peanut, cashew pistachio
4	M	27	W, H	W = mod H = weak	asthma	cashew, brazil, coconut
5	F	54	W, H	H = weak	AR, asthma	cashew, pecans
7	F	30	W, H	W = mod H = strong	AD, AR	peanut, cashew
8	F	43	W, H	W = strong	AD, AR, asthma	peanut, unspecified tree nuts
9	F	35	W	W = strong	AD, AR, asthma	cashew, pecans, almond
10	F	31	W, H	H = strong	AR, asthma	cashew, sunflower
14	F	39	W, H	W = strong	asthma	unspecified tree nuts
20	F	48	W, H	W = weak H = weak	AD, AR asthma	peanut, cashew, chestnut
32	M	38	W, H	W = mod	AD, asthma	cashew, pecan, pistachio
44	F	50	W, H	W = strong H = strong	AD, AR asthma	almond, pecan, pine nut
48	F	62	W	W = weak	none	peanut, cashew, almond
49	F	22	W	W = weak	asthma	pecan, macadamia
81	M	13	W	W = mod	asthma	almond, NE other tree nuts
82	F	24	W	W = weak	none	peanut, walnut, almond, pine nut, NE other tree nuts
83	M	6	W	W = strong	Asthma	walnut, NE other tree nut
91	F	34	W, H	H = strong	AD	cashew
92	M	4	H	H = strong	AD, AR	NE other tree nuts
93	F	11	W, H	H = strong	AD, AR	peanut, NE other tree nuts
94	M	20	H	H = mod	AD, AR	peanut
95	M	2	W, H	H = mod	AD, AR	NE other tree nuts
96	M	10	W, H	H = weak	AD, AR	none reported

^a Patient #s 1–49 correspond to those in previous publications (Wang et al., 2002, 2003; Robotham et al., 2005) and 81–96 represent new patients. Weak = patient IgE bound with weak intensity, mod: moderate, and strong: strong. NE: never eaten.

Table 2
Linear Jug r 4 epitopes and relative IgE-binding signal intensity.

Epitope	AA position	AA sequence	IgE Binding Intensity			
			Pool 1 ^a	Pool 2	Pool 3	Pool 4
1 ^b	1–15	MAKPILLSIYFLIV	++	+	+	++
2	57–71	IESWDPNNQFQCAG	+			
3 HS#1	89–103	YSNAPQLVYIARGRG		+	+	+
4	105–119	TGVLFPGPCPETFEES	++			
5	121–135	RQSQQGQSREFQQDR				+
6	129–143	REFQQDRHQKIRHFR		+	+	+
7	145–159	GDI IAFPAGVAHWSY		+		+
8	201–215	QQQQEYEQHRRQQQR	+	+		+
9	209–223	HRRQQQRQQRPGEHG		+		+
10 ^c HS#2	233–247	VFSGFDADFLADAFN	+++			
11 HS#3	257–271	QSENDHRRSIVRVEG	++		+++	+
12 HS#3	265–279	SIVRVEGRQLQVIRP	+		++	+
13	273–287	QLQVIRPWSREEQE	+	+	+	+
14	281–295	WSREEQEREERKERE	+++	+	+++	++
15	313–327	DDNGLEETICTLRLR		+	++	+
16	377–391	PHWNLNAHSVYALR			+	+
17	385–399	SVVYALRGRAEVQVV				++
18	417–431	LTIPQNFVVKRARN			++	+
19	425–439	VVKRARNEGFEWVSF		++	+++	++
20 HS#4	465–479	LATAFQIPREDARRL	++			+
21	481–495	FNRQESTLVRSRPSR		+		
22	489–503	VRSRPSRSRSSRSER		+		

^aPool 1 included patients 1, 3, 32, and 44; pool 2 included patients 4, 7, 9, and 20; pool 3 included patients 8, 48, and 49; and pool 4 included patients 14, 81, 82, and 83.

^bRed denotes those epitopes bound by sera from all four patient pools.

^cBold print denotes strongly reacting epitopes. HS #1–4 refer to peptides overlapping hot spots indicated in Fig. 1.

3.2. Cor a 9 linear epitopes

Sera from 12 patients with IgE reactive with native and recombinant Cor a 9 (Table 1) (Beyer et al., 2002), were similarly screened for IgE-binding to linear epitopes by probing 63-overlapping solid-phase synthetic Cor a 9 peptides. As presented in Table 3, the 12 serum samples, distributed among three pools, collectively reacted strongly with 3, moderately with 2, and weakly with 20 peptides.

3.3. Primary sequence comparison of linear epitopes of allergenic 11S globulins

To compare the IgE-binding peptides of the walnut (Jug r 4) and hazelnut (Cor a 9) 11S globulins with each other and those previously identified on cashew (Ana o 2) (Wang et al., 2003), peanut (Ara h 3) (Rabjohn et al., 1999; Rouge et al., 2009), and soybean G1 (Gly m 6.01) and G2 (Gly m 6.02) (Zeece et al., 1999; Holzhauser et al., 2009; Helm et al., 2000) 11S globulins, we aligned the sequences and highlighted the corresponding IgE-binding regions (Fig. 1). Many of the epitopes identified in the various allergens were localized to homologous regions. Reactive peptides not showing at least 50% homologous positional overlap with reactive peptides in another allergen were the exception. For ease of discussion, we have highlighted 4 allergenic ‘hot spots’ (HS #1–4, boxed in black dots, Fig. 1) which we define as IgE-binding peptides which displayed positional overlap in 4 or more of the 5 aligned allergens. Those peptides which reacted strongly with at least one of the tree nut-allergic patients’ serum pools are also boxed in blue (Fig. 1).

Most strongly reactive epitope-bearing peptides are in regions also identified as epitopes in one or more other allergens and some are in allergenic ‘hot spots’. For hot spot #3, all compared allergens express reactive IgE-binding peptides in this region, several of which are strongly reactive. Hot spot #2 shows 4 of 6 compared allergens are reactive including strongly reactive peptides for Jug r 4 and Cor a 9 (Tables 2 and 3). Interestingly, the homologous region

for Ana o 2 is not recognized by any of the patient pools even though there is considerable sequence similarity and identity compared to Jug r 4 and Cor a 9.

3.4. Molecular modeling and surface mapping

The 11S globulins are typically homo-hexameric proteins consisting of two trimers assembled in a face-to-face fashion as depicted in Fig. 2. The, IA (hexamer-exposed) and IE (hexamer-occluded) faces of each trimeric subunit refer to those faces of the trimer that are “solvent accessible” and “solvent occluded” when in the hexameric form, respectively. To assess the relationship between solvent exposure and epitope exposure we indicated those residues on the atomic model of the glycinin homo-hexamer (upon which the other models are based) that are solvent exposed in the monomer, trimer and hexamer (≡), the monomer and trimer but occlude in the hexamer (≡), and those only exposed in the monomer (–) in Fig. 1 (labeled “Solv Exposed”). Residues not exposed in the monomer (i.e., are buried) have no symbol and (*) indicates those residues lacking atomic coordinates. Because of the uncertainty of the external loop structure on the modeled tree nut homologues, these were not subjected to surface exposure calculations. Consequently, some residues identified as being on the surface would likely be occluded by the unmodeled loops leading to an overestimation of their abundance.

The data indicate that most epitope-bearing peptides are likely to have a substantial number (≥50%) of exposed residues, at least in the monomer. The patterns of peptide exposure range from a predominance of residues exposed at all higher order configurations (monomer, trimer and hexamer, ≡) to those exposed in the monomer but occluded in the trimer and hexamer (–) and to mixed exposure patterns (–, ≡, and ≡). A minority of epitopes, such as Ana o 2 peptide 137A–N151 and Jug r 4 peptide 145G–Y159, display minimal surface exposure (unmarked residues, ≤35% surface exposed residues). Even one of the hot spots (#1) shows just 8 of 23

Table 3
Linear Cor a 9 epitopes and relative IgE-binding signal intensity.

Epitope	AA position	AA sequence	IgE Binding Intensity		
			Pool 1 ^a	Pool 2	Pool 3
1 ^b	1–15	MAKLILVSFSLCLLV	+	+	+
2	57–71	CQIESWDHNDQQFQC	+		
3 HS#1	89–103	PQYSNAPELIYIERG	+		
4	97–111	LIYIERGRGITGVLF		+	+
5	121–135	PQQSQSQGQRQGGQ	+		+
6	129–143	QRQGGQSQRSEQDR	+		
7	137–151	QRSEQDRHQKIRHFR	++	+	
8	153–167	GDIIALPAGVAHWCY		+	
9	169–183	DGDSPPVTVSLHTN	+		
10	185–199	YANQLDENPRHFYLA	+	+	+
11 ^c	201–215	NPDDEHQRRQGGQFG	+++	+	
12	209–223	QGQQQFGQRRRQQH	++		
13 HS#2	241–255	FSGFDAEFLADAFNV	+++	+++	
14 HS#3	273–287	IVKVEGRLQVVRPER	+		
15 HS#3	281–295	QVVRPERSRQEWERQ	+++	+++	+++
16	289–293	RQEWERQERERESE	+	+	
17	297–311	RQERESEQERERQRR	+	+	
18	305–319	ERERQRRQGGRRDV	+		
19	313–327	GGRGRDVGFEETIC	+		
20	361–375	VLRWLQLSAERGLDQ	+		
21	433–447	RAESEGFVWAFKTN	+		
22 HS#4	473–487	AFQISREEARRLKYN	+	+	
23	481–495	ARRLKYNRQETTLVR		+	
24	489–503	QETTLVRSSSRSSER	+	+	
25	497–511	SRSSSERKRRSESEG	+	+	

^aPool 1 included patients 10, 91, 94, and 96; pool 2 included patients 4, 5, 44, and 93; pool 3 included patient 7, 20, 92, and 95.

^bRed denotes those epitopes bound by sera from all three patient pools.

^cBold print denotes strongly reacting epitopes. HS #1–4 refer to peptides overlapping hot spots indicated in Fig. 1.

residues exposed which are clustered into two separate groups. Linear epitopes are in regions showing alpha helices (e.g., HS #3 and #5) as well as beta turns (e.g., HS #2) (data not shown). Perhaps equally significant, regions that are not identified as expressing linear epitopes also display a range of solvent exposure and secondary structural motifs.

Fig. 3 shows a 3D depiction of the relative positions of the peptides expressing the newly identified IgE-binding epitope-bearing peptides of both Jug r 4 (Fig. 3A–E) and Cor a 9 (Fig. 3F–J). Most regions of the modeled Jug r 4 and Cor a 9 were based on the soybean homo-hexamer crystal structure (depicted as solid surfaces). For regions that were not templated on the atomic structure, we applied energy minimization to assess their likely conformation. We depicted these regions as ribbons to indicate a lower level of structural certainty. Epitopes are highlighted based on the strength of the signals generated and the number of serum pools of binding: epitopes which yielded stronger signals with all pools of patient sera (purple) appear to be localized to the hexamer-occluded faces of Jug r 4 and Cor a 9 (Fig. 3A and F), whereas those reacting strongly by only one pool of patient sera (red) are localized to the hexamer-exposed faces (Fig. 3B and G). Those peptides that reacted moderately or weakly by all pools of patient sera (orange) and that reacted weakly to moderately by at least one but not all pools of patient sera (yellow) are distributed somewhat diffusely about the trimer surfaces (Fig. 3A, B, F and G).

To assess the epitope distribution from the perspective of the monomer, three orientations are shown for each of the two allergens (Fig. 3C–E and H–J). The data indicate that epitopes with varying degrees of binding are found on all faces (exposed and occluded) of the legumin hexamers, trimers, and monomers. Although primary aa sequence homology was high among the compared legumin sequences (Table 4), the conformations of some surface features were found to vary. We highlighted the identified

epitope-bearing peptides of Jug r 4 (Table 2), Cor a 9 (Table 3), Ana o 2 (Wang et al., 2003), and Ara h 3 (Rabjohn et al., 1999), on the monomeric subunits on both the trimer-exposed (top row, Fig. 4) and trimer-occluded (bottom row, Fig. 4) faces of the molecules and color-keyed them according to the degree of pooled patient IgE reactivity as explained above. In general, the epitopes mapped to a variety of locations scattered over the monomeric subunit. Here again, the distribution appeared independent of solvent exposure in the trimeric or hexameric form.

To determine if the mapped linear epitopes correspond to structural motifs common to the 11S globulin allergens, those epitope-containing peptides that occupy homologous positions among the legumin allergens, i.e., allergenic 'hot spots' indicated in Fig. 1, were highlighted and coded (Fig. 5) using a color scheme unrelated to that used in Figs. 3 and 4. Similarity of charge distribution within the epitope 'hot spots' was also assessed (Fig. 6). The 'hot spots' in Fig. 6 were circumscribed in colors coordinated with the highlighted peptides in Fig. 5 to assist in comparison.

For the four mapped epitope hot spots, several structural and biochemical similarities and differences among the tree nut allergens can be noted. HS #4 (blue, Figs. 5 and 6), for example, is situated mainly on the trimer-exposed face, is IgE-reactive in each of the four modeled legumins and displayed a fairly conserved structure and charge distribution pattern though it is less so for the peanut as compared to the three tree nuts. HS #2 (orange), in contrast to HS #4, is on the trimer-occluded face and showed pronounced primary sequence, structural, and charge distribution similarity between the walnut Jug r 4 and hazelnut Cor a 9. It would be interesting to test for cross-reactivity between these two epitopes. The homologous region (HS #4) on cashew Ana o 2 though similar in sequence surprisingly is not recognized by patient IgE. Patients allergic to peanut Ara h 1 and Gly m 6.1 also have IgE which binds to this region but the sequences of these two legumes dif-

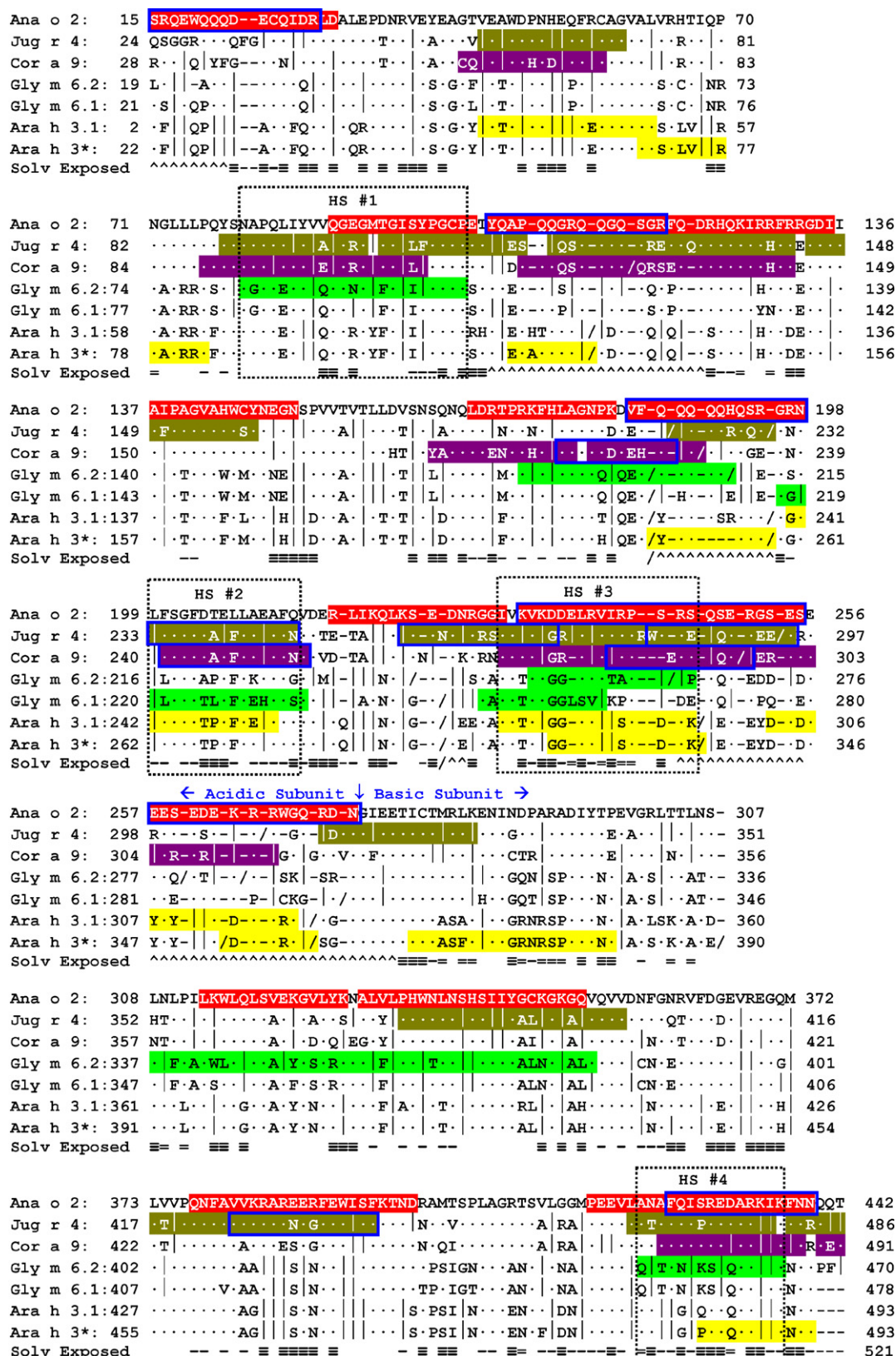


Fig. 1. Amino acid sequence and linear epitope comparison of cashew (Ana o 2), walnut (Jug r 4), hazelnut (Cor a 9), soybean (Gly m 6.2 = Gly m 6.0101 and Gly m 6.2 = Gly m 6.0201), and peanut (Ara h 3) legumins. Segments corresponding to peptides testing positive in IgE SPOTs assays are highlighted. Peptides giving strong signals in the newly reported Jug r 4 and Cor a 9 maps are boxed in blue. Biochemically similar residues are indicated with "|", identical amino acids with ".", positions where a space was added to enhance alignment with "-", and sequences deleted to maximize alignment with "/". Numbered dotted black boxes represent allergenic "hot spots" (HS). Those residues solvent-exposed in the monomer, trimer and hexamer (==), the monomer and trimer but occluded in the hexamer (=) and those only exposed in the monomer (—) of the modeled core structure of the Gly m1 homo-hexamers (PDB 1od5) are indicated in the "Solv Exposed" line. Residues not exposed in the monomer (i.e., are buried) have no symbol. Those residues for which there are no atomic coordinates are also indicated ("."). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

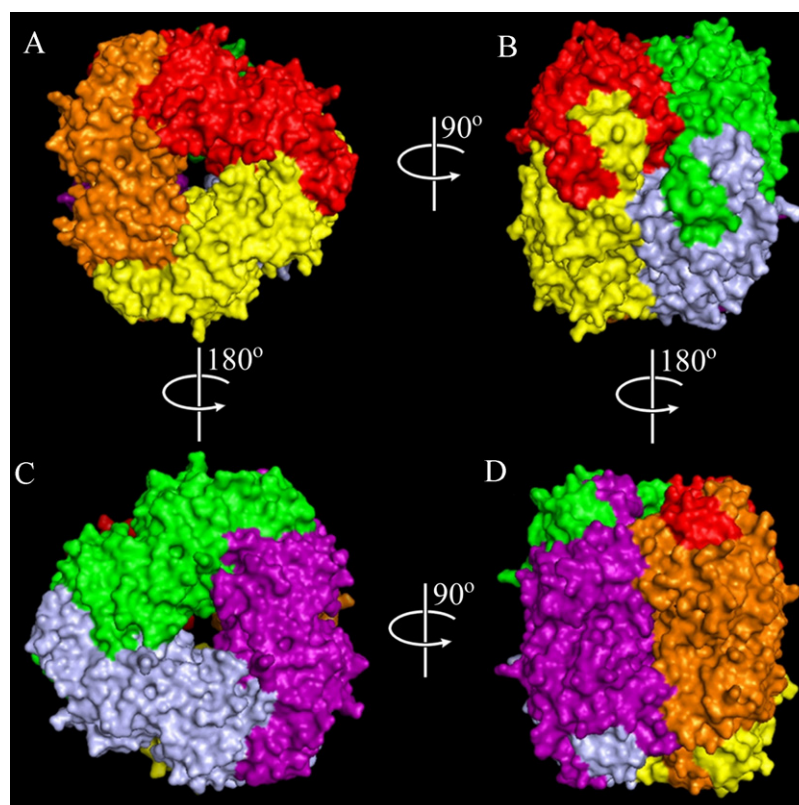


Fig. 2. Atomic structure of soybean glycinin A3B4 homohexamamer, PDB 1od5, showing face-to-face (IE-to-IE) interaction between two trimers as viewed from the planar solvent-accessible IA and solvent-occluded IE faces (A and C) and sides (B and D) of the molecule. The six identical monomeric subunits are individually colored for clarity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

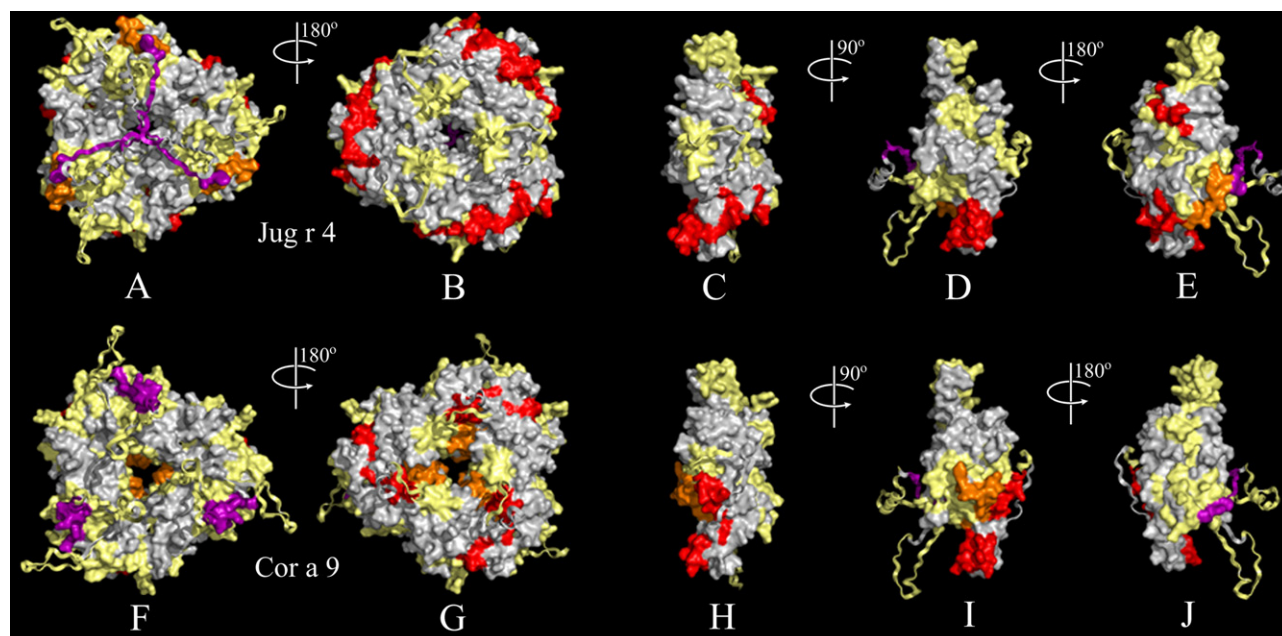


Fig. 3. Homology modeling and reactive peptide surface mapping of the Jug r 4 and Cor a 9 homotrimers and monomer subunits. Epitopes are highlighted based on the strength and frequency of pooled sera binding: yellow, weak to moderate binding by some pools and no binding by at least one pool; orange, binding with similar strength by all pools; red, strong binding by at least one pool; purple, strong binding by all pools. (A) Jug r 4 trimeric IE, also termed 'occluded', and (B) IA, also termed 'accessible', faces. (C) Monomer oriented as in (B), (D) rotated to show the trimer solvent-occluded face and (D) rotated to show the trimer solvent-accessible face (i.e., the faces that are occluded/exposed when the subunits are in the trimeric state). (F–J) Corresponding Cor a 9 models oriented as described above. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 4

Amino acid sequence comparison of allergenic legumins.

	Gly m 6.0201	Ana o 2	Jug r 4	Cor a 9	Ara h 3.1
Glycinin	100%	45/65 (474) ^a	46/64 (489)	45/64 (499)	56/72 (510)
Ana o 2		100%	57/73 (478)	53/69 (494)	42/59 (485)
Jug r 4			100%	73/84 (494)	45/60 (484)
Cor a 9				100%	45/60 (493)
Ara h 3					100%

^a Percent identity/similarity. The number of amino acids included in each alignment is shown in parentheses.

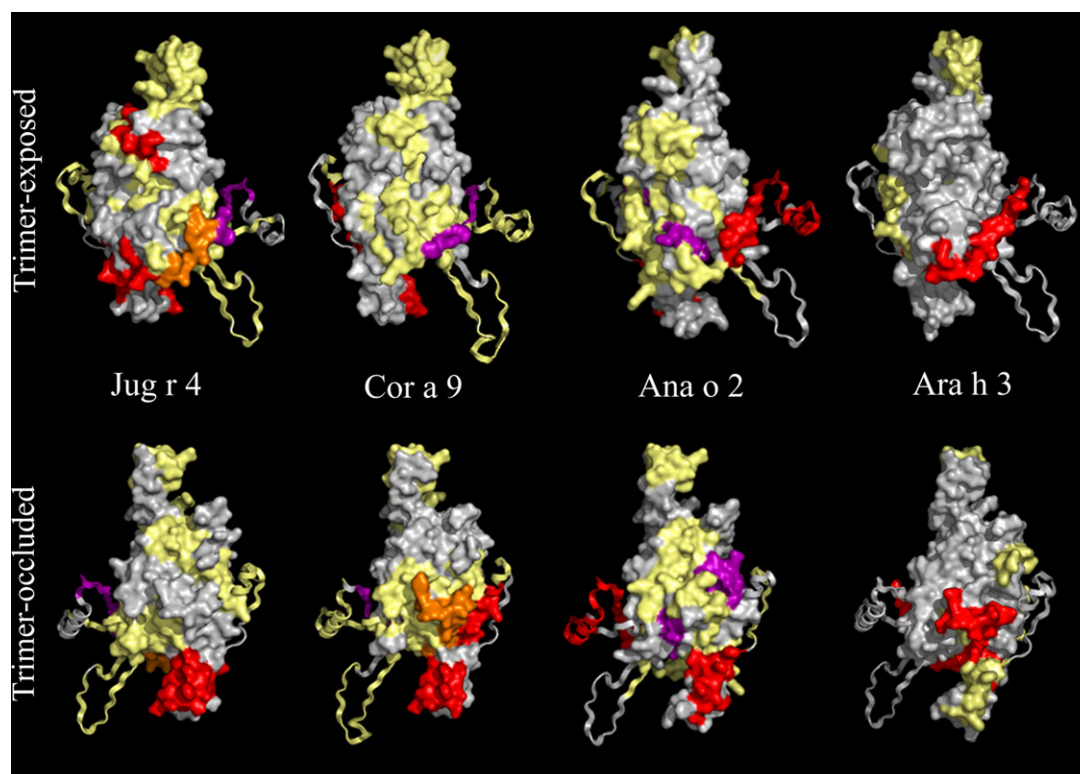


Fig. 4. Comparative homology modeling and reactive peptide surface mapping of Jug r 4, Cor a 9, Ana o 2, and Ara h 3 monomers. Epitopes highlighted based on pooled patient reactivity as in Fig. 3. The trimer solvent-accessible faces shown in the top row and the trimer solvent-occluded faces on the bottom. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

fer significantly from those of the tree nuts as does the modeled structure and charge distribution (for Ara h 3). HS #3 (burgundy) represents a region targeted by IgE on all four modeled legumins. Again Jug r 4 and Cor a 9 share considerable sequence, structural and charge similarity whereas Ana o 2 and, to an even greater extent, Ara h 3 differ. It is worth noting that 3 of the 4 hot spots (#2–4) are comprised primarily of solvent-exposed residues, at least at the level of the free monomer (see Fig. 1). HS #1 (green), in contrast, is largely buried within the monomer. It is also worth noting that none of the hot spots were coincident with the untemplated loops even though such loops would be expected to represent protruding surface features readily accessible for antibody recognition.

4. Discussion

The structural properties which distinguish those proteins that are allergenic in some foods but are typically non-allergenic in other foods and the structural bases for food allergen cross-reactivity are incompletely understood (Aalberse et al., 2001; Jenkins et al., 2005; Roggen, 2006; Breiteneder and Mills, 2006; Radauer and Breiteneder, 2006; Untersmayr and Jensen-Jarolim, 2006). We have selected the 11S globulins as a model system to further investi-

gate these issues, in part, because several allergenic legumins have already been analyzed for their structure and IgE epitope distribution (Stanley et al., 1997; Zeece et al., 1999; Helm et al., 2000; Adachi et al., 2001; Beyer et al., 2002; Teuber et al., 2003a; Wang et al., 2003; Barre et al., 2007; Rouge et al., 2009). Here we report linear epitope maps of 11S globulins in two widely consumed allergenic tree nuts, walnut and hazelnut, and compare the epitope maps at the primary sequence level and in 3D molecular models to each other and to those previously described for cashew, peanut, and soybean.

Taken together, 11S globulins of the analyzed tree nuts and legumes display from ~8 to 15 linear epitope-bearing peptide regions that are scattered throughout the length of the acidic and basic subunits (Fig. 1). The newly identified epitopes on Jug r 4 and Cor a 9 display varying degrees of patient IgE reactivity (Tables 2 and 3). For most tree nut 11S globulin epitopes, there seems to be no correlation between either the breadth, or intensity of patient IgE reactivity and the location of the epitopes with respect to accessible or occluded faces of the Jug r 4 or Cor a 9 hexamers or trimers (Figs. 1 and 3, Tables 2 and 3).

A comparison of the epitope distribution among the compared nut and seed 11S globulins reveals four regions most frequently recognized by patient IgE (Fig. 1). We termed these common IgE-

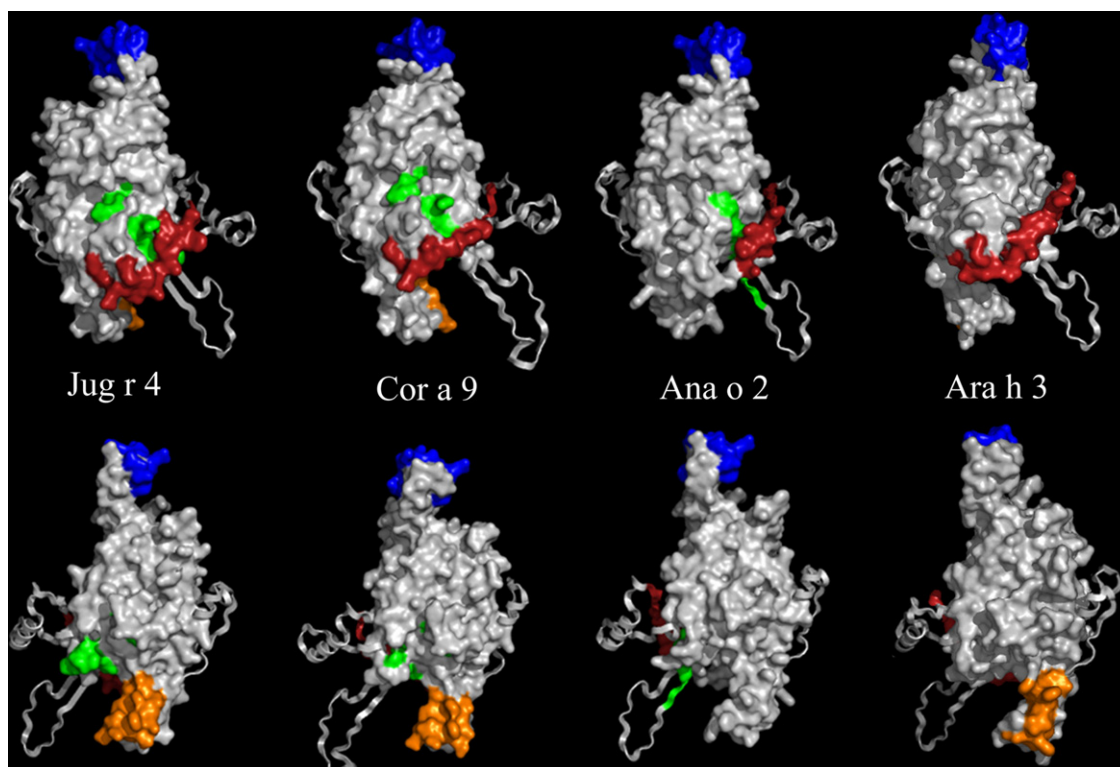


Fig. 5. Topological comparison of allergenic “hot spots” on the surface of Jug r 4, Cor a 9, Ana o 2, and Ara h 3. Hot spot (HS, #'s correspond to Fig. 1): #1, green; #2, orange; #3, burgundy; #4, blue. Orientations as in Fig. 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

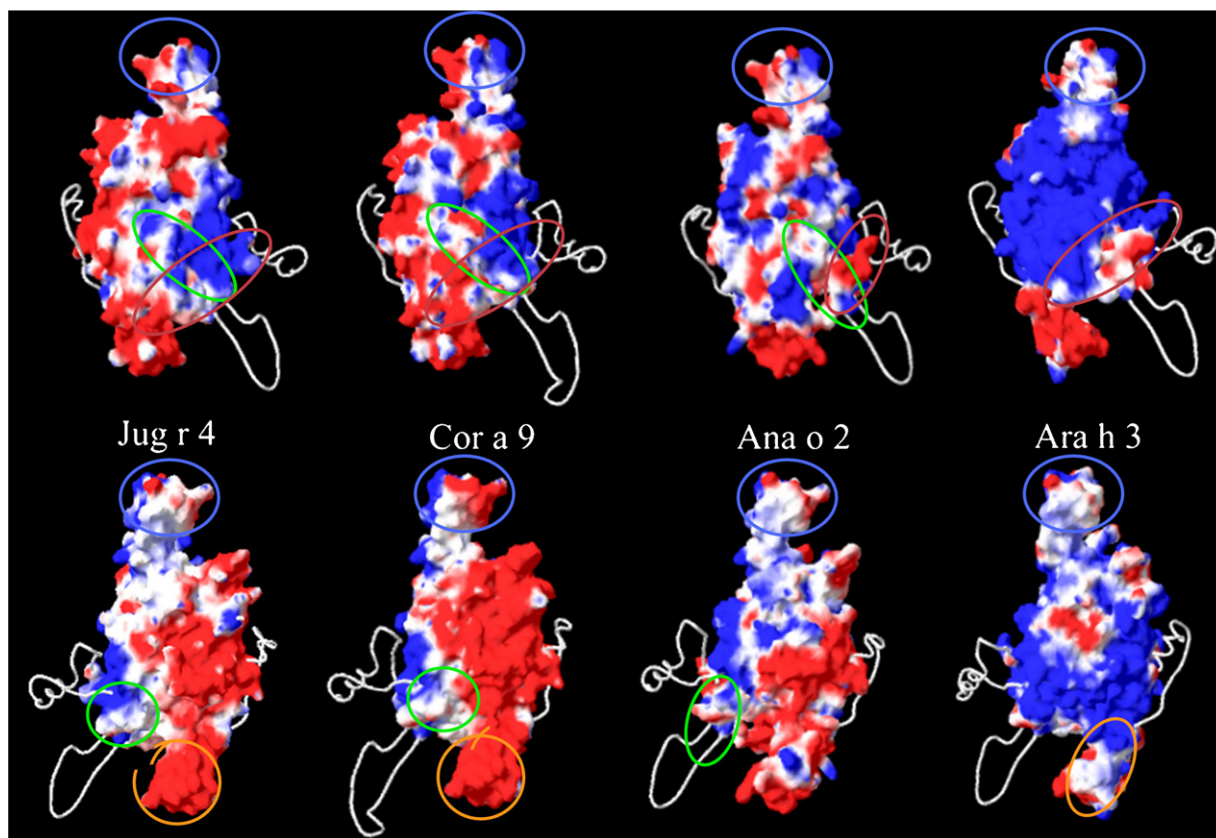


Fig. 6. Comparison of surface charge distribution maps of Jug r 4, Cor a 9, Ana o 2, and Ara h 3. Red = negatively charged, blue = positively charged, and white = neutral regions. Ovals circumscribe IgE-binding “hot spots” and are color-coded as depicted in Fig. 5. Orientations as in Figs. 4 and 5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

binding epitopes, 'hot spots', and mapped the antigenic surfaces onto our 3D models (Fig. 5). HS #2 (orange) and #4 (blue), for example, share primary sequence similarity between Jug r 4 and Cor a 9 (100% for HS #2 and 86% for HS #4) and appear structurally related as evidenced by similarities in the surface topology maps (Fig. 5) and corresponding charge distribution maps (Fig. 6). HS #1 (green) was identified as an epitope on all three tree nuts, and soybean but not the peanut, even though soybean and peanut share considerable homology in this region. The surface charge distribution of this region was similar for walnut and hazelnut but less so for cashew (Fig. 6). An interesting aspect of this hot spot is that the majority of the residues are buried in the interior suggesting that denaturation and/or fragmentation of the monomeric subunit is needed to expose the epitope and that the resultant IgE antibodies would probably not be reactive with intact undenatured monomeric or higher order molecular forms.

Barre et al. (2007) predicted that HS #3 would be a common epitopic region among the tree nut allergens based on their analysis of the previously identified IgE-binding epitopes of Ana o 2 (Wang et al., 2003) and Ara h 3 (Rabjohn et al., 1999) and our data bear this out.

The shared structural homology in HS #2 (orange, Fig. 5) and HS #4 (blue, Fig. 5) represents a likely candidate for the observed cross-reactivity among some tree nut allergens (Wallowitz et al., 2006). A comparison of the peanut Ara h 3 hot spots to those of the tree nuts revealed that peanut-allergic patients recognized 3 of the 4 hot spot regions (HS #2–4, Fig. 1). However, homologous regions in the peanut differ significantly from those of the tree nuts in sequence and charge with the exception of HS #4 in which the peanut and hazelnut shares 86% similarity. Rouge et al. used SPOTs analysis to identify eight Ara h 3 epitopes (Rouge et al., 2009). They too noted that the homologous sequence for most peanut epitopes were divergent from the tree nuts but singled out two regions (their peptides #1, 49ALSR-LRRP63 and #7, 337ICTA-IYNP357) with conserved sequence and predicted that these might account for peanut-tree nut cross-reactivity. Interestingly, with one partial exception, neither homologous region was recognized in either the tree nuts or even in the more closely related soybean sequence. Together these data may help explain the low incidence of reported clinical cross-reactivity between tree nuts and peanut, especially if similar patterns of discordant epitope sharing are observed for other in-common allergenic proteins (Teuber et al., 2003b).

It should be noted that the various highlighted segments represent epitope-containing peptides and not the actual epitope contact residues. Because the boundaries of the linear epitope-bearing peptides screened with patients' sera varied between allergens, the color-coded hot spot patches depicted in Fig. 5, for example, would not necessarily match up even if the actual contact residues were homologous, i.e., differences in the position and extent of the highlighted peptides should not be construed as evidence for structural differences between epitopes. Also, a lack of recognition may reflect the allergenic status of the particular patient population sampled rather than an intrinsic inability to elicit an IgE response—other patient sera may yet be identified which recognize these regions.

The distribution pattern of the identified linear epitopes on (in) an allergenic food protein may provide clues to the degree of its *in vivo* digestion, denaturation, and cellular processing prior to its exposure to B cells (Untersmayr and Jensen-Jarolim, 2006). The structural complexity of the typical legumin presents an interesting opportunity to gauge the structural state of the protein as it interacts with the humoral immune system. As a dimer of homotrimers, each succeeding smaller molecular form (hexamer, trimer, monomer, acidic and basic monomeric subunits, or peptide fragments thereof) displays more solvent- and thus antibody-accessible surface. It has been reported that for soybean G2 glycinin, most linear epitopes were localized to the hexamer-exposed (IA) face of the trimer sug-

gesting that the hexamers may remain intact upon interaction with the humoral immune system (Helm et al., 2000). In contrast, the broad distribution of the cognate IgE-binding epitopes on tree nut and peanut 11S globulins in both exposed and cryptic sites that we present suggests a significant degree of processing by the gut, or perhaps by antigen presenting cells, prior to immune recognition, an area that needs further investigation. Complicating our interpretation of the data to some degree is the fact that maturational processing (i.e., proteolytic cleavage) of 11S globulins into forms that hexamerize within plants is incomplete resulting in residual unprocessed preprotein that remains in the trimeric configuration in nut/seed tissues prior to consumption (Shewry et al., 2003). Also, because soybean has 5 genes encoding 11S globulin proteins (Nielsen et al., 1989), other, as yet undescribed isoforms may exist in peanut and the tree nuts.

Various other molecular forms are present in the nut/seed tissues as indicated by multiple products of varying molecular weight observed upon Western blot analysis of nut/seed extract (Teuber et al., 2002). Also, surface exposure, *per se*, does not appear to ensure IgE recognition as several such regions (e.g., the segments homologous to Ana o 2 aa 356–372) are devoid of epitopes for all of the nuts and legume seeds compared.

The question is thus raised as to whether or not epitopes are being assayed in biologically relevant molecular forms. For example, an allergen may be largely digested, denatured or fragmented in the induction phase of the immune response but remain in a more native configuration at the site of mast cell triggering. Thus, an identified linear epitope may be strongly reactive when assayed with patients' sera but concealed within an allergen in its native configuration. Knowledge of the allergen's structure as it is presented to the effector arm immune system is essential when considering which epitopes will have the greatest potential for contributing to a clinical reaction. Such information is generally lacking for most systems (Untersmayr and Jensen-Jarolim, 2006). Because patient gut-associated mast cells and tissue IgE profiles cannot be readily assayed, this question will be difficult to address. Perhaps animal models can be exploited in this regard (Untersmayr and Jensen-Jarolim, 2006).

Complete food allergens (capable of both sensitization and symptom elicitation) have been considered resistant to digestion by the gut (Astwood et al., 1996; Bredehorst and David, 2001; Taylor and Hefle, 2001). A recent report indicates that peanut and soybean 11S globulins are readily digested by pepsin at low pH (van Boxel et al., 2008) despite predictions to the contrary (Rouge et al., 2009). However, susceptibility to digestion does not preclude the 11S globulins from serving as sensitizers as a number of other examples of readily digestible allergens capable of directly sensitizing patients have been reported (Untersmayr and Jensen-Jarolim, 2006). It is not known if cashew, hazelnut and walnut 11S globulins are similarly labile but, if so, this might explain the widespread topological distribution of their linear IgE-reactive epitopes including the hexamer- and trimer-occluded surfaces as well as those largely internal peptides of the monomer.

It is important to note that we and others have only mapped the linear nut epitopes primarily because of technical difficulty in identifying conformational epitopes (Roggen, 2006; Gershoni et al., 2007; Oezguen et al., 2008). Although conformational epitopes are sometimes considered of greater significance in allergic conditions as compared to linear epitopes, there are situations where the reverse appears to be true. For example, early responses to milk allergens are primarily to conformational epitopes but later responses are to linear epitopes (Untersmayr and Jensen-Jarolim, 2006). In other examples of food allergy, conformational epitopes are generally associated with mild reactions and linear epitopes with more intense persistent allergy (Beyer, 2003; Sicherer and Sampson, 2006). Because of the extent to which the legumin sur-

faces that we have mapped are comprised of linear epitopes, it is likely that there is considerable overlap with any conformational epitope footprints.

Defining common structural features among specific cross-reactive epitopes may enhance our ability to predict, *a priori*, whether homologous regions on other molecules of interest are likely to be similarly recognized and could inform as to the mutations that might best reduce allergenicity in, for example, genetically engineered allergen homologues developed for allergy vaccine application (Roggen, 2006; Gershoni et al., 2007). A broad compilation of examples of both cross-reactive and non-cross-reactive epitopes from a variety of sources might eventually reveal molecular motifs which favor the elicitation of IgE antibody should such motifs exist.

Acknowledgements

Funded in part by USDA CSREES 2003-01212 (KHR, SKS) and the Food Allergy Initiative (New York, New York) (SST).

References

- Aalberse, R., Akkerdaas, J., van Ree, R., 2001. Cross-reactivity of IgE antibodies to allergens. *Allergy* 56, 478–490.
- Adachi, M., Kanamori, J., Masuda, T., Yagasaki, K., Kitamura, K., Mikami, B., Utsumi, S., 2003. Crystal structure of soybean 11S globulin: glycinin A3B4 homohexamer. *Proc. Natl. Acad. Sci.* 100, 7395–7400.
- Adachi, M., Takenaka, Y., Gidamis, A.B., Mikami, B., Utsumi, S., 2001. Crystal structure of soybean proglycinin A1aB1b homotrimer. *J. Mol. Biol.* 305, 291–305.
- Asero, R., Mistrello, G., Roncarolo, D., Amato, S., 2004. Walnut-induced anaphylaxis with cross-reactivity to hazelnut and brazil nut. *J. Allergy Clin. Immunol.* 358–360.
- Astwood, J.D., Leach, J.N., Fuchs, R.L., 1996. Stability of food allergens to digestion in vitro. *Nat. Biotechnol.* 14, 1269–1273.
- Barre, A., Jacquet, G., Sordet, C., Culerrier, R., Rouge, P., 2007. Homology modelling and conformational analysis of IgE-binding epitopes of Ara h 3 and other legumin allergens with a cupin fold from tree nuts. *Mol. Immunol.* 44, 3243–3255.
- Beyer, K., 2003. Characterization of allergenic food proteins for improved diagnostic methods. *Curr. Opin. Allergy Clin. Immunol.* 3, 189–197.
- Beyer, K., Grishina, G., Bardina, L., Grishin, A., Sampson, H.A., 2002. Identification of an 11S globulin as a major hazelnut food allergen in hazelnut-induced systemic reactions. *J. Allergy Clin. Immunol.* 110, 517–523.
- Bock, S.A., Munoz-Furlong, A., Sampson, H.A., 2001. Fatalities due to anaphylactic reactions to foods. *J. Allergy Clin. Immunol.* 107, 191–193.
- Bredehorst, R., David, K., 2001. What establishes a protein as an allergen? *J. Chromatogr. B: Biomed. Sci. Appl.* 756, 33–40.
- Breiteneder, H., Mills, C., 2006. Structural bioinformatic approaches to understand cross-reactivity. *Mol. Nutr. Food Res.* 50, 628–632.
- Breiteneder, H., Radauer, C., 2004. A classification of plant food allergens. *J. Allergy Clin. Immunol.* 113, 821–830.
- de Leon, M.P., Drew, A.C., Glaspole, I.N., Suphioglu, C., Rolland, J.M., O'Hehir, R.E., 2005. Functional analysis of cross-reactive immunoglobulin E antibodies: peanut-specific immunoglobulin E sensitizes basophils to tree nut allergens. *Clin. Exp. Immunol.* 35, 1056–1064.
- 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.
- Gershoni, J.M., Roitburd-Berman, A., Siman-Tov, D.D., Tarnovitski Freund, N., Weiss, Y., 2007. Epitope mapping: the first step in developing epitope-based vaccines. *BioDrugs* 21, 145–156.
- Goetz, D.W., Whisman, B.A., Goetz, A.D., 2005. Cross-reactivity among edible nuts: double immunodiffusion crossed immunoelectrophoresis, and human specific IgE serologic surveys. *Ann. Allergy Asthma Immunol.* 95, 45–52.
- Helm, R.M., Cockrell, G., Connaughton, C., Sampson, H.A., Bannon, G.A., Beilinson, V., Nielsen, N.C., Burks, A.W., 2000. A soybean G2 glycinin allergen. 2. Epitope mapping and three-dimensional modeling. *Int. Arch. Allergy Immunol.* 123, 213–219.
- Holzhauser, T., Wackermann, O., Ballmer-Weber, B.K., Bindslev-Jensen, C., Scibilia, J., Perono-Garoffo, L., Utsumi, S., Poulsen, L.K., Vieths, S., 2009. Soybean (*Glycine max*) allergy in Europe: Gly m 5 (beta-conglycinin) and Gly m 6 (glycinin) are potential diagnostic markers for severe allergic reactions to soy. *J. Allergy Clin. Immunol.* 123, 452–458.
- Jenkins, J.A., Griffiths-Jones, S., Shewry, P.R., Breiteneder, H., Mills, E.N.C., 2005. Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: an in silico analysis. *J. Allergy Clin. Immunol.* 115, 163–170.
- Nielsen, N.C., Dickinson, C.D., Cho, T.J., Thanh, V.H., Scallion, B.J., Fischer, R.L., Sims, T.L., Drews, G.N., Goldberg, R.B., 1989. Characterization of the glycinin gene family. *Plant Cell* 1, 313–328.
- Oezgüen, 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.
- Poltronieri, P., Cappello, M.S., Dohmae, N., Conti, A., Fortunato, D., Pastorello, E.A., Ortolani, C., Zacheo, G., 2002. Identification and characterisation of the IgE-binding proteins 2S albumin and Conglutin gamma in almond (*Prunus dulcis*) seeds. *Int. Arch. Allergy Immunol.* 128, 97–104.
- Rabjohn, P., Helm, E.M., Stanley, J.S., West, C.M., Sampson, H.A., Burks, A.W., Bannon, G.A., 1999. Molecular cloning and epitope analysis of the peanut allergen Ara h 3. *J. Clin. Invest.* 103, 535–542.
- Radauer, C., Breiteneder, H., 2006. Pollen allergens are restricted to few protein families and show distinct patterns of species distribution. *J. Allergy Clin. Immunol.* 117, 141–147.
- Robotham, J.M., Wang, F., Seamon, V., Teuber, S.S., Sathe, S.K., Sampson, H.A., Beyer, K., Seavy, M., Roux, K.H., 2005. Ana o 3, an important cashew nut (*Anacardium occidentale* L.) allergen of the 2S albumin family. *J. Allergy Clin. Immunol.* 115, 1284–1290.
- Roggen, E.L., 2006. Recent developments with B-cell epitope identification for predictive studies. *J. Immunotoxicol.* 3, 137–149.
- Rouge, P., Culerrier, R., Sabatier, V., Granier, C., Rance, F., Barre, A., 2009. Mapping and conformational analysis of IgE-binding epitopic regions on the molecular surface of the major Ara h 3 legumin allergen of peanut (*Arachis hypogaea*). *Mol. Immunol.* 46, 1067–1075.
- Schwede, T., Kopp, J., Guex, N., Peitsch, M.C., 2003. Swiss-model: an automated protein homology-modeling server. *Nucleic Acids Res.* 31, 3381–3385.
- Shewry, P.R., Jenkins, J.A., Beaudoin, F., Mills, E.N., 2003. The classification, functions and evolutionary relationships of plant proteins in relation to food allergies. In: Mills, E.N., Shewry, P.R. (Eds.), *Plant Food Allergens*. Blackwell Publishing Ltd., Malden, MA, pp. 24–41.
- 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 random digit dial telephone survey: a 5-year follow-up study. *J. Allergy Clin. Immunol.* 112, 1203–1207.
- Sicherer, S.H., Sampson, H.A., 2006. Food allergy. *J. Allergy Clin. Immunol.* 117, S470–S475.
- Stanley, J.S., King, N., Burks, A.W., Huang, S.K., Sampson, H., Cockrell, G., Helm, R.M., West, C.M., Bannon, G.A., 1997. Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. *Arch. Biochem. Biophys.* 342, 244–253.
- Taylor, S.L., Hefle, S.L., 2001. Will genetically modified foods be allergenic? *J. Allergy Clin. Immunol.* 107, 765–771.
- Teuber, S.S., Comstock, S.S., Sathe, S.K., Roux, K.H., 2003a. Tree nut allergy. *Curr. Allergy Asthma Rep.* 3, 54–61.
- Teuber, S.S., Sathe, S.K., Peterson, W.R., Roux, K.H., 2002. Characterization of the soluble allergenic proteins of cashew nut (*Anacardium occidentale* L.). *Agric. Food Chem.* 50, 6543–6549.
- Teuber, S., Peterson, W., Uratsu, S., Dandekar, A., Roux, K., Sathe, S., 2003b. Identification and cloning of Jug r 4, a major food allergen from English walnut belonging to the legumin group. *J. Allergy Clin. Immunol.* 111, S248.
- Theobald, D.L., Wuttke, D.S., 2008. Accurate structural correlations from maximum likelihood superpositions. *PLoS Comput. Biol.* 4, e43.
- Theobald, D.L., Wuttke, D.S., 2006. THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* 22, 2171–2172.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Untersmayr, E., Jensen-Jarolim, E., 2006. Mechanisms of type I food allergy. *Pharmacol. Ther.* 112, 787–798.
- van Bostel, E.L., van den Broek, L.A.M., Koppelman, S.J., Gruppen, H., 2008. Legumin allergens from peanuts and soybeans: effects of denaturation and aggregation on allergenicity. *Mol. Nutr. Food Res.* 52, 674–682.
- van Gunsteren, W.F., Berendsen, H.J., 1987. Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry. *J. Comput. Aided Mol. Des.* 1, 171–176.
- Wallowitz, M., Peterson, W.R., Uratsu, S., Cornstock, S.S., Dandekar, A.M., Teuber, S.S., 2006. Jug r 4, a legumin group food allergen from walnut (*Juglans regia* Cv. Chandler). *J. Agric. Food Chem.* 54, 8369–8375.
- Wallowitz, M., Chen, R.J.Y., Tzen, J.T.C., Teuber, S.S., 2007. Ses i 6, the sesame 11S globulin, can activate basophils and shows cross-reactivity with walnut in vitro. *Clin. Exp. Allergy* 37, 929–938.
- Wang, F., Robotham, J.M., Teuber, S.S., Sathe, S.K., Roux, K.H., 2003. Ana o 2, a major cashew nut (*Anacardium occidentale* L.) allergen of the legumin family. *Int. Arch. Allergy Immunol.* 132, 27–39.
- Wang, F., Teuber, S.S., Robotham, J.M., Tawde, P., Sathe, S.K., Roux, K.H., 2002. Ana o 1, a cashew (*Anacardium occidentale*) allergen of the vicilin seed storage protein family. *J. Allergy Clin. Immunol.* 110, 160–166.
- Zeece, M.G., Beardslee, T.A., Markwell, J.J.P., Sarath, G., 1999. Identification of an IgE-binding region in soybean acidic glycinin G1. *Food Agric. Immunol.* 11, 83–90.