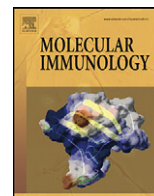




Contents lists available at ScienceDirect

Molecular Immunology

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



Crystal structure of Ara h 3, a major allergen in peanut

Tengchuan Jin^a, Feng Guo^a, Yu-wei Chen^a, Andrew Howard^a, Yu-Zhu Zhang^{a,b,*}

^a Department of Biological, Chemical, and Physical Sciences, Illinois Institute of Technology, Chicago, IL 60616, United States

^b National Center for Food Safety and Technology, Illinois Institute of Technology, 6502 South Archer Road, Summit-Argo, IL 60501, United States

ARTICLE INFO

Article history:

Received 21 November 2008

Received in revised form 15 January 2009

Accepted 25 January 2009

Available online xxx

Keywords:

Peanut allergy

IgE epitope mapping

Allergen stability

Immunotherapeutic agent design

X-ray crystallography

ABSTRACT

The prevalence of food allergy has increased dramatically in recent years. Tremendous research progress has been made in understanding the pathophysiological mechanisms of allergy and in identifying and characterizing food allergens. Peanut is a major food allergen source and Ara h 3 is a major peanut allergen. Using overlapping short peptides, several linear IgE-binding epitopes in Ara h 3 have been defined before. However, the structure of Ara h 3 of the native allergen is not clear and information on conformational epitopes is lacking. Structural characterization of allergens is required for understanding the allergenicity of food allergens and for the development of immunotherapeutic agents. Previously, we have reported the crystallization of Ara h 3 purified from raw peanut. Here we report the crystal structure of Ara h 3 at 1.73 Å resolution. Mapping of the previously defined linear epitopes on the crystal structure of Ara h 3 indicated that linear epitopes with more solvent exposure were those indicated by the literature to react with more patient sera. The structure of Ara h 3 may be used to assess the importance of conformational epitopes in further investigations.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Food allergy has become a major health concern in industrialized countries in the past two decades (Branum and Lukacs, 2008; Sampson, 2004; Wang and Sampson, 2007). Most food allergies are triggered by the immunoglobulin E (IgE) recognition of food allergens (Asero et al., 2007; Nowak-Węgrzyn, 2007; Ozol and Mete, 2008). To date, there is no cure for food allergies although immunotherapy using engineered hypoallergenic proteins is considered promising for treating the disease (de Leon et al., 2007b; Li et al., 2003; Pons et al., 2005; Sampson, 2004). Peanut and tree nuts count for most of the fatal and near fatal food allergy cases in the United States (Sampson et al., 1992; Yunginger et al., 1988). Cross-reaction of peanut allergic subjects to other tree nuts, legumes, and oil seeds (Beardslee et al., 2000; de Leon et al., 2007a, 2005; Goetz et al., 2005; Wensing et al., 2003) poses additional risk to peanut allergic patients. Hence, extensive research has been devoted to peanut allergy.

To date, 11 peanut allergens (Ara h 1–11) have been officially recognized by the Allergen Nomenclature Sub-committee of the International Union of Immunological Societies (<http://www.allergen.org/Allergen.aspx>, accessed Nov. 14, 2008). Stability and resistance to protease digestion are among the factors that con-

tribute to the allergenicity of food allergens. Ara h 2 and Ara h 3 have been reported to have activities similar to bovine pancreatic trypsin inhibitor (BPTI) (Dodo et al., 2004; Maleki et al., 2003) and a peanut trypsin inhibitor with high degree of homology to part of the N-terminal domain of Ara h 3 has been isolated (Dodo et al., 2004). Recombinant peanut allergens have been expressed and many linear IgE-binding epitopes on the major allergens have been reported (Burks et al., 1997; Rabjohn et al., 1999; Stanley et al., 1997). However, structural information of peanut allergens is not available, except for Ara h 6 whose protease resistant core has been studied by high resolution NMR (Lehmann et al., 2006).

The peanut major allergen Ara h 3 is an 11S seed storage protein. It is known that the 11S protein has multiple isoforms in numerous species. Each isoform of the 11S globulin is generally encoded by a single gene (Wang et al., 2002) that produces a precursor that is post-translationally cleaved by an asparaginyl endopeptidase after the formation of an interchain disulfide bond between the N-terminal and the C-terminal subunits (Albillos et al., 2008; Badley et al., 1975; Kitamura et al., 1976). The orthologs of Ara h 3 in many nuts and legumes are also known to be allergens. The 11S storage protein from soybean is known to be a hexameric globulin that is composed of various mixtures of 5 different isoforms (Badley et al., 1975). Five different genes were known to encode the isoforms of Ara h 3 in peanut (Yan et al., 2005). The purification and crystallization of these allergens from natural sources is considered difficult because of heterogeneity (Adachi et al., 2003). Nevertheless, to obtain accurate structural information on Ara h 3, we purified one isoform of Ara h 3 from dry peanut kernels and obtained diffraction quality crystals for high resolution X-ray structure determination

* Corresponding author at: Department of Biological, Chemical, and Physical Sciences, Illinois Institute of Technology, Chicago, IL 60616, United States.
Tel.: +1 312 567 3484; fax: +1 312 567 3494.

E-mail address: zhangy@iit.edu (Y.-Z. Zhang).

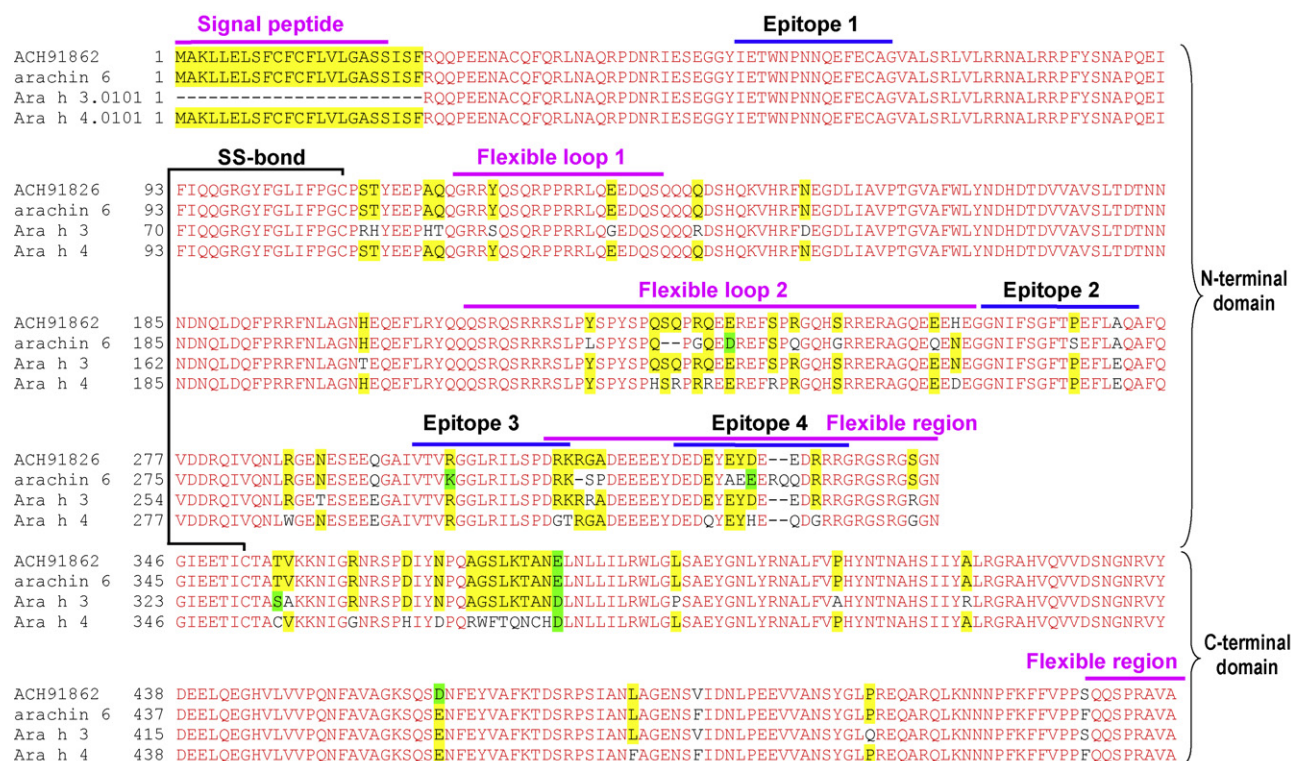


Fig. 1. Multiple sequence alignment of Ara h 3. Ara h 3 isolated in this study (ACH91862), arachin 6 (ABL14270), and the officially recognized peanut allergen Ara h 3.0101 (AAC63045), and Ara h 4.0101 (AAD47382) sequences were aligned using ClustalW. Identical residues are shown in red; conserved residues are shaded in yellow; and similar residues are shaded in green. The N-terminus of the C-terminal domain starts a new line and the N- and C-terminal domains are indicated at the right side of the sequence alignment. The positions of the flexible regions as well as previously defined IgE epitopes (Rabjohn et al., 1999) are indicated on the top of the sequences. The disulfide bond link between the N- and C-terminal domains is also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

(Jin et al., 2007). Here, we report the crystal structure of Ara h 3 determined at 1.73 Å resolution and the mapping of previously reported linear IgE-binding epitopes in the structure.

2. Materials and methods

2.1. Genomic DNA isolation

Genomic DNA was isolated for cloning Ara h 3 as our unpublished data on isolating the 11S genes of a number of other species from mRNA indicated that many of them do not contain introns. 100 mg of peanut embryos were separated from dried raw peanuts. Genomic DNA was prepared from the embryos with the RNeasy Plus Mini Kit and QIAprep Spin Miniprep Kit (QIAGEN, Valencia, CA) using a procedure combining the manufacturer's protocols for

Table 1

Structure refinement statistics.

Refinement	
Resolution range (Å)	50.0–1.73
Number of reflections	
Working set	49,465
Test set ^a	2631
Number of atoms	
Protein	3291
Water	285
Sodium ion	1
R/R _{free} (%)	18.57/24.83
rmsd from ideal geometry	
Bond lengths (Å)	0.026
Bond angles (degree)	2.10

^a 5% of the total reflections randomly chose.

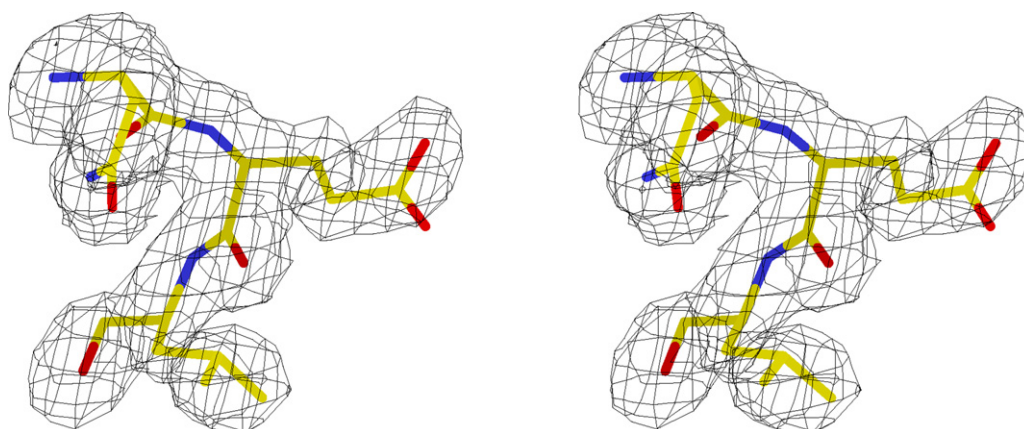


Fig. 2. Identifying the isoform of purified Ara h 3 based on electron density. A stereo view of a shake-and-omit Fo-Fc electron-density map contoured at 3.5σ calculated with E381, which aligned with an Asp residue in Ara h 3.0101 and Ara h 4.0101, and its neighboring residues, N379 and L382, omitted.

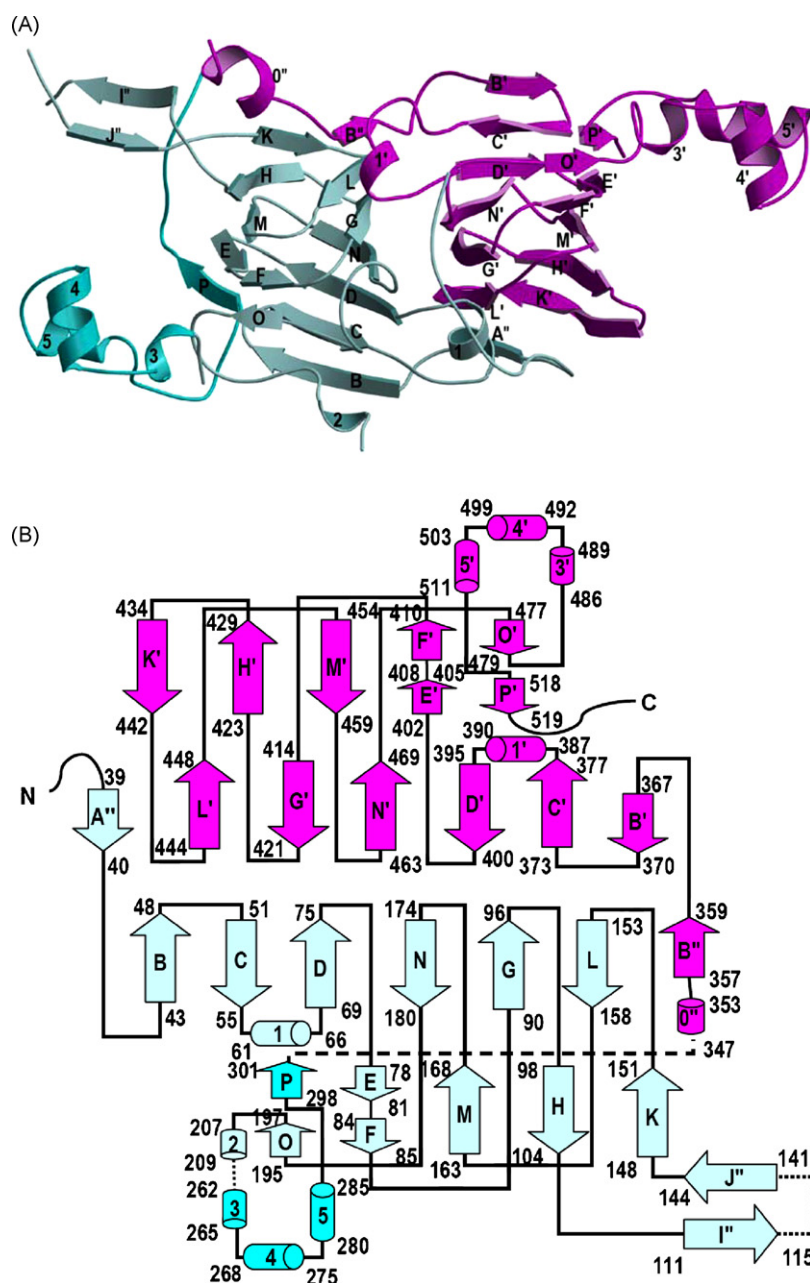


Fig. 3. Structure of Ara h 3. (A) Unit cell structure of Ara h 3. The N-terminal domain is shown in cyan (with its peanut trypsin inhibitor homologous region lightened) and the C-terminal domain is shown in magenta. Strands are labeled alphabetically, and helices are numbered. Element labels in the C-terminal domain are primed. The first strand in the N-terminal peptide is part of a β -sheet of the C-terminal domain and is double-primed. Similarly, the first strand in the C-terminal peptide is double-primed. Strand I and J in the N-terminal domain are also double primed because they would become part of a β -sheet of the C-terminal of another protomer in the hexameric native allergen. (B) A schematic diagram shows the N- and C-terminal domain cupin fold. The coloring and labeling scheme are the same as in A. Dotted lines indicate un-located flexible loops and the broken line indicates that the peptide connection between the N- and C-terminal domains was cleaved. The residue positions of the starts and ends of the helices and strands are indicated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the two kits. Briefly, the embryos were placed in a mortar containing enough liquid nitrogen to keep the sample submerged and ground with a pestle. After grinding, the powder and the liquid nitrogen were transferred to an RNase-free, DNase-free, liquid-nitrogen-cooled 2 ml microcentrifuge tube. As soon as the liquid nitrogen was evaporated, 450 μ l of Buffer RLT was added to the sample. The sample was then vigorously vortexed and the amount of the sample corresponding to 50 mg of embryos was transferred to a QIAshredder spin column (purchased separately from QIAGEN) and centrifuged for 2 min at 14,000 rpm in a microcentrifuge. The supernatant of the flow-through was transferred to a QIAprep spin column placed in a 2 ml collection tube and spun for 2 min

at 14,000 rpm. Column washing and DNA elution were carried out following the miniprep kit protocol including the usage of the optional steps. Fifty microliters of Milli-Q water was used to elute the genomic DNA.

2.2. PCR and cloning

Judged by the electron-density map, Ara h 3 purified in this study is very similar to arachin 6 (ABL14270). Thus, two primers (ctcgaag-gagatctatacc ATGGCTAAGCTTCTTGAGCTTCTTTTTC, forward, and tgggtctcagggccgcg TTAAGCCACAGCCCTCGGAGAC, reverse) were designed to amplify the coding sequence of this Ara h 3 isoform.

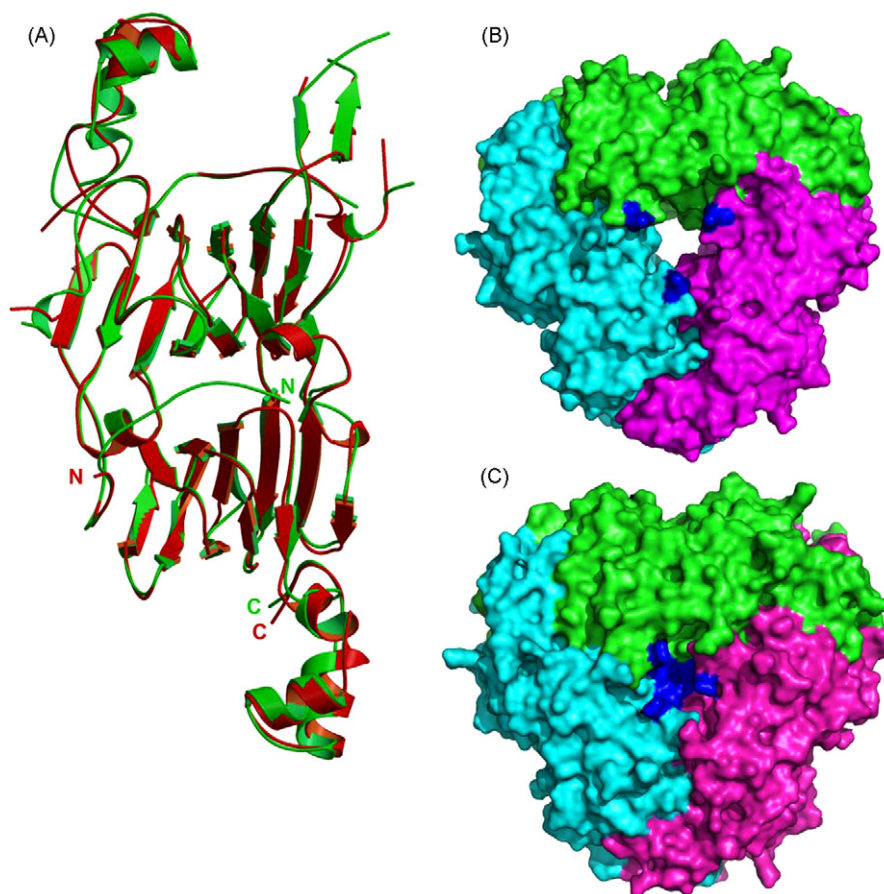


Fig. 4. Comparison between soybean glycinin and Ara h 3. (A) Structure superposition of soybean glycinin (PDB:1OD5A) and Ara h 3 (PDB:3C3V). Ara h 3 is shown in green and soybean glycinin is shown in red. The N- and C-termini of the two structures are shown in corresponding colors. (B) Molecular surface of soybean glycinin trimer. (C) Molecular surface of Ara h 3 trimer. The monomers in both trimers are shown in magenta, cyan, and green, respectively. The N-terminus of the glycinin structure (N7) and the Ara h 3 N-terminal residues up to the residue corresponding to soybean glycinin-N7 by structure alignment are shown in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

The lower case portions of the primers were added as priming site for a second PCR. The PCR product was separated on a 1% agarose gel and purified to serve as template in the second PCR with primers ggaatggggacaagttgtacaaaaagcaggctcgaaggagatctataccatg (forward) and ggaatgggggaccactttgtacaagaaagctgggtctcagggccgc (reverse) to add the attachment sites for GatewayTM (Invitrogen) cloning. After separation and purification, the PCR product was used in an *in vitro* site specific recombination with the pDONR/221 donor vector using BP enzyme mix (Invitrogen) to create pD-arah3. After *E. coli* transformation, positive colonies were used to prepare pD-arah3 for sequencing in order to determine the coding sequence for mature Ara h 3. The cDNA sequence of one of the isolated isoforms of the Ara h 3 gene which encodes the Ara h 3 purified in this study has been submitted to the GenBank and the accession numbers is FJ041908.

2.3. Structure refinement

Ara h 3 was purified from raw peanut and crystallized by the vapor-diffusion method. X-ray data were collected at the SER-CAT 22ID beamline at the Advanced Photon Source (APS), Argonne National Laboratory. Data processing with HKL2000 (Otwinowski and Minor, 1997) and Phaser (McCoy et al., 2005; Storoni et al., 2004) molecular-replacement results were reported previously (Jin et al., 2007).

Structure refinement was carried out with simulated annealing using CNS (Brunger et al., 1998) followed by refinement with

REFMAC5 (Vagin et al., 2004) implemented in CCP4i (Potterton et al., 2003). The refinement was alternated with model building and model improvement using coot (Emsley and Cowtan, 2004).

The final structure was refined with all data to 1.73 Å resolution and the final model was checked by PROCHECK (Laskowski et al., 1993) and Molprobit Validation (Lovell et al., 2003). The quality of the structure model was also checked with a shake-and-omit protocol by first introducing random errors up to 0.3 Å to the coordinates of the final refined structure using the program pdsset distributed with CCP4 (Potterton et al., 2003). For each region to be checked, the concerned region of the shaken structure was manually omitted and 20 cycles of restrained refinements were carried out using REFMAC5 (Vagin et al., 2004). This was followed by inspecting the Fo-Fc map together with the final refined structure. Solvent accessible surface of the final structure was calculated using the program NACCESS (Hubbard and Thornton, 1993), with a probe that has a radius of 1.4 Å. The surface of each amino acid type in the ala-X-ala tripeptide was used as reference to calculate the percentage of residue surface that is exposed on the surface of the monomeric, trimeric, and hexameric structure of Ara h 3. The structure of Ara h 3 has been submitted to the Protein Data Bank under accession number 3C3V.

2.4. Structure analyses

Structure based sequence alignment and the corresponding superposition of structures were carried out using the program

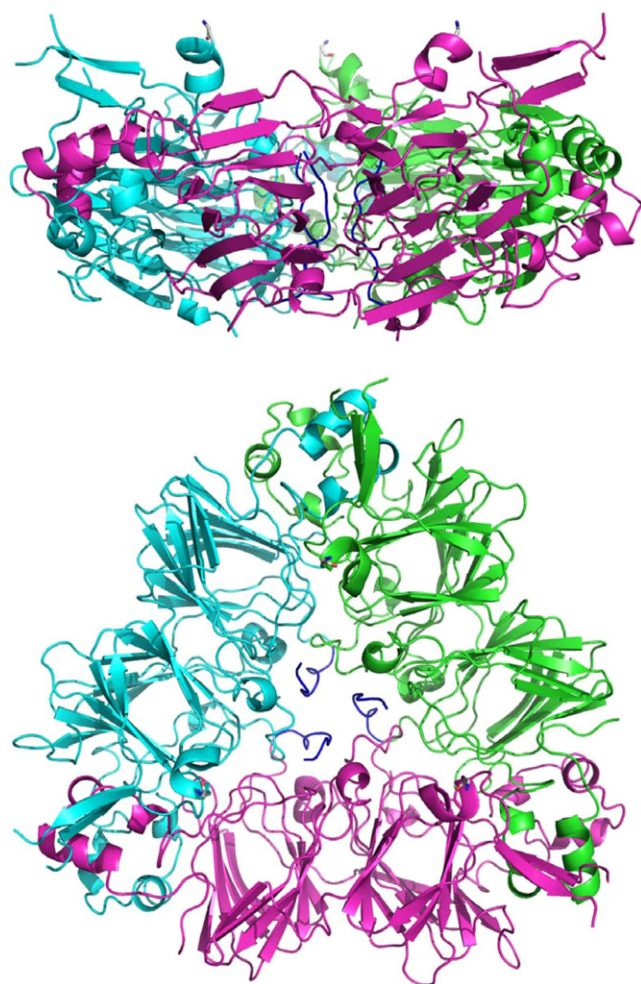


Fig. 5. Burial of the N-terminus of the C-terminal domain. The trimeric structure of Ara h 3 is represented by a ribbon diagram in the upper panel, with the three monomers shown in magenta, green, and cyan, respectively. The N-terminal residue (G346) of the C-terminal domain of Ara h 3 is represented as sticks using the CPK coloring scheme. The lower panel shows the Ara h 3 in the same representation as in the upper panel after the structure was rotated 90° about a horizontal axis pointing to the right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Mustang (Konagurthu et al., 2006). Residue wise solvent accessible surface (SAS) was calculated using the program Naccess (Hubbard and Thornton, 1993). Molecular graphics were prepared using the programs rasmol (Sayle and Milner-White, 1995), molscript (Kraulis, 1991), Raster3D (Merritt and Murphy, 1994), and Pymol (<http://pymol.org/>).

3. Results

3.1. Isolation of Ara h 3 coding sequences from peanuts

The purification of Ara h 3 from natural source and its crystallization has been described previously (Jin et al., 2007). However, there are at least 5 isoforms of Ara h 3 (Yan et al., 2005), and 12 different Ara h 3 sequences were in the NCBI protein database, possibly from different cultivars. To determine the exact sequence of the Ara h 3 purified in this study, all the available Ara h 3 sequences were referenced to make the side chains best fit the 2Fo-Fc map during the initial rounds of model building. This allowed us to match all the modeled peptides to one of the Ara h 3 sequences in the database (ABL14270:arachin 6) with a few exceptions. To resolve these uncertainties and to ensure the right sequence is used

for the final structure refinement, we decided to isolate the Ara h 3 genes from the same bag of peanuts that was used for purifying the protein. Using primers derived from the arachin 6 gene (GenBank: EF080819), two Ara h 3 coding sequences were isolated from genomic DNA of this peanut variant. The peptide sequence deduced from one of the isolated Ara h 3 cDNA fit the entire electron map, except for the un-located residues in the flexible loops and for residues with un-locatable side chain atoms. The sequence of this isoform of Ara h 3 is shown in Fig. 1. The sequence identity between this Ara h 3 and arachin 6, Ara h 3.0101, and Ara h 4.0101 are 90%, 95%, and 93%, respectively. Without the regions that could not be located in the electron-density map, these sequence identities are 98%, 95%, and 95%, respectively. An Ara h 3 gene with four introns has been reported (GenBank: AF510854) (Viquez et al., 2004). The sequence identity between Ara h 3 isolated in this study and that deduced from the Ara h 3 gene with introns (Dodo et al., 2004) is 88%. Information on the expression profiles of the different Ara h 3 genes during seed development remains to be obtained.

3.2. The structure of Ara h 3

The final refined model gave R/R_{free} values of 18.6%/24.8% for all data to 1.73 Å (Fig. 2 and Table 1). The rmsd from ideal empirical values are 0.026 Å for bond lengths and 2.1° for bond angles with no main chain bond length deviates more than 0.05 Å and two main chain bond angles deviate (10.1° and 10.5°) more than 10° from "ideal" small molecule values. On the Ramachandran plot with Molprobity Validation (Lovell et al., 2003), 95.9% of all residues are in favored regions and all residues are in allowed regions. In the final refined structure, there is one Ara h 3 molecule in the asymmetric unit. It includes 398 amino acids, 285 water molecules, and one sodium ion.

The overall structure of Ara h 3 consists of an N-terminal domain and a C-terminal domain, each assumes a cupin fold and they are related by a pseudo-dyad axis (Fig. 3A). Three regions of the N-terminal domain (G119–Q138 and Q212–G259, and D311–N345) and the tail of C-terminal domain (S522–A530) were disordered (Figs. 1 and 3B). For the side chains of residues N201, D278, and N363, only the β -carbon could be located and included in the final structure and for residues D462 and E504, only the β - and γ -carbons of their side chains were located and included.

3.3. Comparison with soybean glycinin

The crystal structure of the 11S glycinin from a genetically modified soybean variant, in which all but the A3B4 glycinin isoforms were deleted, has been reported (Adachi et al., 2003). Structure alignment between Ara h 3 and the soybean 11S glycinin (PDB:1OD5A) is shown in Fig. 4A. In general, the structures of these two proteins are largely superimposable. The greatest differences are seen in the loops and in the helical regions. Structure based alignment of the two sequences of the structurally ordered parts of Ara h 3 and soybean glycinin resulted in a sequence identity of 47.2%. Gaps in the alignment account for 5.3%. Over 50% of the gap positions are not the same as those in the result of a BLAST alignment of the regions of the two proteins.

Similar to soybean glycinin A3B4, mature Ara h 3 is a hexamer formed by a head-to-head association of two trimers. One of the notable differences between the packing of the two structures is near the three-fold symmetric axis of the trimer. While there is a large cavity channel in the doughnut shaped soybean glycinin (Fig. 4B), there is no channel present in the Ara h 3 trimer (Fig. 4C) because an N-terminal peptide of mature Ara h 3 folded back along the three-fold axis of the trimer (Fig. 5).

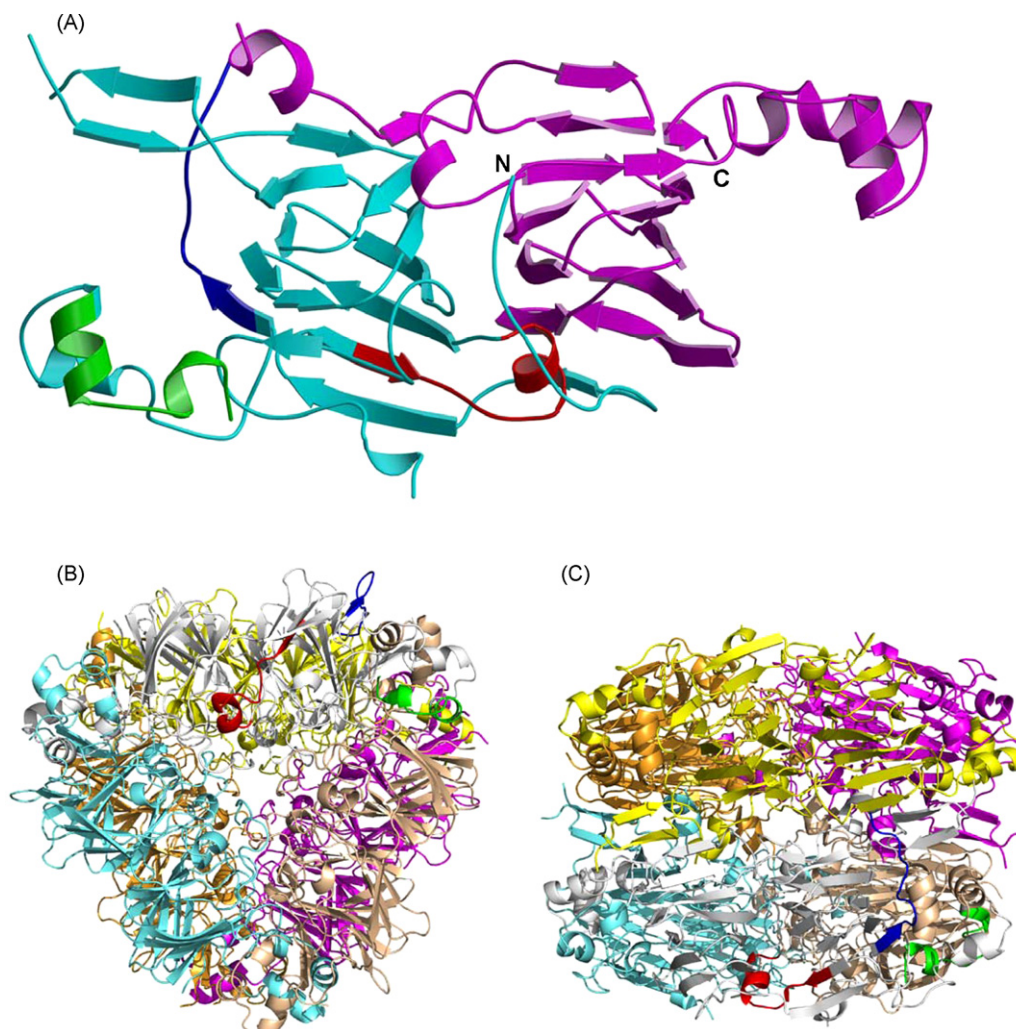


Fig. 6. Mapping of previously reported linear IgE epitopes. (A) A ribbon diagram of an Ara h 3 monomer. The N-terminal domain is shown in cyan (but with linear epitope 1, 2, and 3 shown in red, green, and blue, respectively) and the C-terminal domain is shown in magenta. (B) Hexameric Ara h 3 is represented with each of the monomers shown in a different color. In the gray monomer, linear epitope 1, 2, and 3 are shown in red, green, and blue, respectively. (C) The hexamer represented in an orthogonal orientation as in B after it was rotated 90° about a horizontal axis pointing to the right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

3.4. The Ara h 3 hexamer

The 11S seed storage protein is synthesized as a single chain, but a peptide bond between the N-terminal and C-terminal domains is cleaved in the mature protein (Dodo et al., 2004; Muntz, 1998). Same as in soybean glycinin (Adachi et al., 2003), the N-terminus of the C-terminal domain is located at the trimer–trimer interface (Fig. 5) and there was no space left for additional residues at its N-terminal, suggesting that the peptide linkage between the two domains would prevent the formation of an Ara h 3 hexamer.

3.5. Mapping of known linear IgE epitopes

Four Ara h 3 linear epitopes (epitope 1–4) have been identified in a study that screened 15-mer overlapping peptides derived from the Ara h 3 sequence using 8 patient sera (Rabjohn et al., 1999). We sought to map these previously defined linear epitopes on the structure of Ara h 3 determined in this study. Epitope 4 is located in the C-terminal region of the N-terminal domain of Ara h 3. It is disordered in the crystal and not included in the refined structure (Figs. 1 and 3B). The last three residues of epitope 3 were also not located in the structure. Linear epitope 1 and 2 as well as the structured portion of epitope 3 was mapped on the final

1.73 Å monomeric structure of Ara h 3 (Fig. 6A) and in the Ara h 3 homohexamer (Fig. 6B and C). These epitopes are partially exposed on the surface of the native allergen (Fig. 7). The side chains of most of the residues identified as critical for IgE-binding in a previous study (Rabjohn et al., 1999), however, are completely or nearly completely buried. For instance, P58 and N59 in epitope 1 were the only residues determined to be critical for IgE-binding. P58 is largely buried and inaccessible for IgE binding in the native allergen (Table 2). Critical IgE-binding residues in epitope 2 (F263, F266, F270, and L271) are all available for IgE-binding if Ara h 3 is presented as a monomer. In the native hexameric Ara h 3, however, these residues are buried (Table 2). L305, I307, L308, and P310 were determined to be the critical residues in epitope 3 (Rabjohn et al., 1999). The average SAS for L305, I307, and L308 is 12%. The SAS for P310 is 61%, but this value would have been lower if it were not the last locatable residue of the N-terminal domain of Ara h 3. The reference surface was calculated for the residue in the middle of an ala-X-ala tripeptide, not at the end of a peptide.

4. Discussion

The purification and crystallization of a homohexameric Ara h 3 from peanuts may indicate the feasibility of obtaining three-

Table 2
Solvent accessible surfaces (SAS) of the amino acids in the linear IgE epitopes.

		AA	I53	E54	T55	W56	N57	<u>P58</u>	<u>N59</u>	N60	Q61	E62	F63	E64	C65	A66	G67
Epi1	monomer, trimer, and hexamer	SAS(Å ²)	6.37	9.68	0.9	1.91	34.4	15.0	78.6	55.0	58.6	6.78	1.94	71.6	0.55	0.13	4.63
		%SAS	3.6	5.6	0.1	0.8	23.9	11.0	54.6	38.2	32.8	3.9	1.0	41.6	0.4	0.1	5.8
		AA	G260	N261	I262	<u>F263</u>	S264	G265	<u>F266</u>	T267	P268	E269	<u>F270</u>	<u>L271</u>	<u>A272</u>	Q273	A274
Epi2	monomer	SAS(Å ²)	78.0	8.86	62.9	39.6	32.7	69.9	98.8	76.2	50.9	119.7	122.6	6.91	36.1	135.9	76.32
		%SAS	97.4	6.2	35.9	19.8	28.1	87.3	49.5	54.7	37.4	69.5	61.5	3.9	33.5	76.1	70.7
	trimer & hexamer	SAS(Å ²)	42.7	0.69	1.92	0	25.68	3.15	0	25.19	50.85	91.66	0.16	0	36.12	26.02	0
		%SAS	53.3	0.5	1.1	0	22	3.9	0	18.1	37.4	53.2	0.1	0	33.5	14.6	0
		AA	V299	T300	V301	R302	G303	G304	<u>L305</u>	R306	<u>I307</u>	<u>L308</u>	S309	<u>P310</u>	D311	R312	K313
Epi3	monomer	SAS(Å ²)	21.76	58.74	31.02	240.1	53.77	10.86	52.64	192.3	51.34	43.03	38.85	166.2			
		%SAS	14.4	42.2	20.5	100.6	67.1	13.6	29.5	80.5	29.3	24.1	33.3	122.1			
	trimer	SAS(Å ²)	6.96	56.83	3.52	238.0	53.77	9.7	16.78	156.5	30.26	43.03	38.85	166.2			
		%SAS	4.6	40.8	2.3	99.7	67.1	12.1	9.4	65.5	17.3	24.1	33.3	122.1			
	hexamer	SAS(Å ²)	6.96	56.83	3.52	238.0	53.77	9.7	16.78	156.5	30.26	18.31	23.08	83.42			
		%SAS	4.6	40.8	2.3	99.7	67.1	12.1	9.4	65.5	17.3	10.2	19.8	61.3			

Underlined bold face letters represent critical residues for IgE binding (Rabjohn et al., 1999). Residue A272 in epitope 2 is shaded because it is not identical in the sequence of Ara h 3 isolated in this study and that in (Rabjohn et al., 1999). The last three residues in epitope 3 were disordered in the crystal and not included in the solvent accessible surface calculations. Trimer formation only affects the SAS in epitope 2 and 3 and trimer–trimer association only affects epitope 3.

dimensional structures of the 11S seed storage protein from other species, although it is previously thought to be very difficult to isolate the 11S protein to high purity for crystallographic studies due to the complication of their molecular heterogeneity (Adachi et al., 2003). Structural information on additional 11S allergens may facilitate the understanding of the structural basis of their allergenicity and their possible allergenic cross-reactivity.

Immunotherapy using engineered hypoallergenic proteins is promising to become an effective treatment of food allergies (Li et al., 2003; Sampson, 2004), and IgE epitope mapping of food allergens is a prerequisite for engineering hypoallergenic immunotherapeutic agents (Shreffler et al., 2004). In 1999, a study with 18 peanut allergic patient sera showed that recombinant Ara h 3 purified under denaturing conditions was recognized by IgE molecules in 8 of the sera in a Western blot analysis (Rabjohn et al., 1999). The same study screened 15-mer overlapping peptides derived from Ara h 3 sequence with these 8 patient sera and identified 4 Ara h 3 linear epitopes (epitope 1–4), all of which are in the N-terminal domain (Rabjohn et al., 1999). In a recent study with 77 patient sera (Shreffler et al., 2004), recombinant Ara h 3 arrayed on microscope slides was recognized by 77% of the sera. The microarray experiments also tested 20-mer overlapping peptides covering the N-terminal domain of Ara h 3 and did not identify any additional linear major epitopes (Shreffler et al., 2004).

Because the peptide linkage between the N- and C-terminal domains of Ara h 3 would prevent the formation of an Ara h 3 hexamer, the trimer-to-hexamer transition is only possible after the cleavage of the peptide bond between the two domains and the dis-

placement of the loop that linked the two domains. This indicates an obstacle one has to overcome in expressing native recombinant Ara h 3 and the 11S allergens from other species. Simply inserting the coding sequence of an 11S protein into an expression vector may not produce a recombinant form of the protein in its native hexameric state in an expression host. Consistent with this, the recombinant expression of soybean proglycinin in bacteria produced a trimeric protein (Adachi et al., 2001). Thus, previous attempts to identify Ara h 3 recognizing sera using recombinant Ara h 3 would have missed at least those sera that contained only IgE antibodies that bind to conformational epitopes include residues that is also part of the trimer–trimer interface, regardless whether the recombinant protein is separated on a denaturing SDS gel (Rabjohn et al., 1999) or spotted on a microarray chip (Shreffler et al., 2004).

That the previously defined linear epitopes are partially exposed on the surface of the native allergen (Fig. 7) may suggest that they are available for IgE binding even if the intact protein is presented as an allergen. On the other hand, the burial of critical residues of the linear epitopes in the Ara h 3 structure may suggest that native Ara h 3 is most likely not to be recognized by the IgE antibodies that recognized linear epitope 1 and 2 in the 8 patient sera. This may indicate that the allergen is degraded as a result of digestion and the linear epitopes are exposed to interaction with the immune system. However, allergens initiate allergic reactions by cross-linking IgE molecules on the surface of the mast cells and basophils (Galli et al., 2008). Because a short linear epitope with critical residues close to one another is unlikely to associate with more than one IgE molecule, a larger fragment of the allergen or the intact native aller-

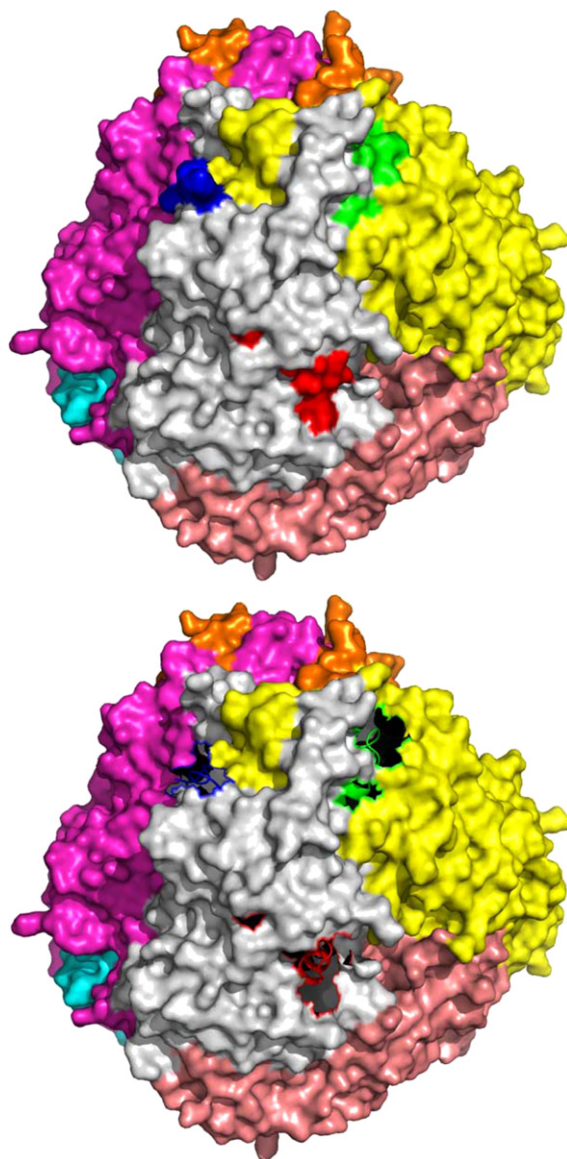


Fig. 7. Solvent exposure of previously reported linear IgE epitopes. The molecular surface of the hexameric Ara h 3 is shown with the surfaces of each of the monomers shown in a different color (upper panel). Epitope 1, 2, and 3 in the gray monomer are shown in red, green, and blue, respectively. In the lower panel, the hexameric Ara h 3 is shown the same as in the upper panel, but with the epitope 1, 2, and 3 shown in the ribbons representation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

gen that encompasses at least two linear epitopes should be much more effective in stimulating allergic reactions. Epitope 4 and part of epitope 3 are flexible in the native protein. IgE antibodies that recognized these epitopes are likely to recognize Ara h 3 in its native state. Because the native allergen fully exposes six copies of epitope 4, this epitope may be more effective in stimulating allergic reactions than the other epitopes if the intact protein is to interact with the immune system. It was reported that the number of patient sera recognized epitope 3 and epitope 4 (Rabjohn et al., 1999; Shreffler et al., 2004) is significantly larger than the number of sera recognized epitope 1 and 2. This may suggest that the intact peanut allergen can be absorbed and interact with the immune system. In fact, it is known that a portion of the ingested food antigens is absorbed and transported throughout the body in an immunologically intact form, even through the guts of the normal subjects (Husby et al., 1987; Sampson, 2004). Thus, whether conformational IgE epitopes

play a more important role than currently known in persistent food allergies, such as peanut allergy, needs further investigation.

Acknowledgements

This work was supported by a fund from Illinois Institute of Technology. X-ray diffraction data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source (APS), Argonne National Laboratory. Use of the APS was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract W-31-109-Eng-38.

References

- 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. U.S.A.* 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.
- Albillos, S.M., Jin, T., Howard, A., Zhang, Y., Kothary, M.H., Fu, T.J., 2008. Purification, crystallization and preliminary X-ray characterization of prunin-1, a major component of the almond (*Prunus dulcis*) allergen amandin. *J. Agric. Food Chem.* 56, 5352–5358.
- Asero, R., Ballmer-Weber, B.K., Beyer, K., Conti, A., Dubakiene, R., Fernandez-Rivas, M., Hoffmann-Sommergruber, K., Lidholm, J., Mustakov, T., Oude Elberink, J.N., Pumphrey, R.S., Stahl Skov, P., van Ree, R., Vlieg-Boerstra, B.J., Hiller, R., Hourihane, J.O., Kowalski, M., Papadopoulos, N.G., Wal, J.M., Mills, E.N., Vieths, S., 2007. IgE-mediated food allergy diagnosis: current status and new perspectives. *Mol. Nutr. Food Res.* 51, 135–147.
- Badley, R.A., Atkinson, D., Hauser, H., Oldani, D., Green, J.P., Stubb, J.M., 1975. The structure, physical and chemical properties of the soy bean protein glycinin. *Biochim. Biophys. Acta* 412, 214–228.
- Beardslee, T.A., Zeece, M.G., Sarath, G., Markwell, J.P., 2000. Soybean glycinin G1 acidic chain shares IgE epitopes with peanut allergen Ara h 3. *Int. Arch. Allergy Immunol.* 123, 299–307.
- Branum, A.M., Lukacs, S.L., 2008. Food Allergy Among U.S. Children: Trends in Prevalence and Hospitalizations. NCHS data brief, No 10. U.S. National Center for Health Statistics, Hyattsville, MD.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L., 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Cryst. D54*, 905–921.
- Burks, A.W., Shin, D., Cockrell, G., Stanley, J.S., Helm, R.M., Bannon, G.A., 1997. Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. *Eur. J. Biochem.* 245, 334–339.
- de Leon, M.P., Drew, A.C., Glaspole, I.N., Suphioglu, C., O'Hehir, R.E., Rolland, J.M., 2007a. IgE cross-reactivity between the major peanut allergen Ara h 2 and tree nut allergens. *Mol. Immunol.* 44, 463–471.
- 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. Allergy* 35, 1056–1064.
- de Leon, M.P., Rolland, J.M., O'Hehir, R.E., 2007b. The peanut allergy epidemic: allergen molecular characterisation and prospects for specific therapy. *Expert Rev. Mol. Med.* 9, 1–18.
- Dodo, H.W., Viquez, O.M., Maleki, S.J., Konan, K.N., 2004. cDNA clone of a putative peanut (*Arachis hypogaea* L.) trypsin inhibitor has homology with peanut allergens Ara h 3 and Ara h 4. *J. Agric. Food Chem.* 52, 1404–1409.
- Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Galli, S.J., Tsai, M., Piliponsky, A.M., 2008. The development of allergic inflammation. *Nature* 454, 445–454.
- 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.
- Hubbard, S.J., Thornton, J.M., 1993. NACCESS. Computer Program, Department of Biochemistry and Molecular Biology, University College London.
- Husby, S., Foged, N., Host, A., Svehag, S.E., 1987. Passage of dietary antigens into the blood of children with coeliac disease. Quantification and size distribution of absorbed antigens. *Gut* 28, 1062–1072.
- Jin, T., Howard, A., Zhang, Y.Z., 2007. Purification, crystallization and initial crystallographic characterization of peanut major allergen Ara h 3. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 63, 848–851.
- Kitamura, K., Takagi, T., Shibasaki, K., 1976. Subunit structure of soybean 11S globulin. *Agric. Biol. Chem.* 40, 1837–1844.
- Konagurthu, A.S., Whisstock, J.C., Stuckey, P.J., Lesk, A.M., 2006. MUSTANG: a multiple structural alignment algorithm. *Proteins* 64, 559–574.
- Kraulis, P.J., 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946–950.

- Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283–291.
- 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, X.M., Srivastava, K., Grishin, A., Huang, C.K., Schofield, B., Burks, W., Sampson, H.A., 2003. Persistent protective effect of heat-killed *Escherichia coli* producing “engineered”, recombinant peanut proteins in a murine model of peanut allergy. *J. Allergy Clin. Immunol.* 112, 159–167.
- Lovell, S.C., Davis, I.W., Arendall III, W.B., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* 50, 437–450.
- Maleki, S.J., Viquez, O., Jacks, T., Dodo, H., Champagne, E.T., Chung, S.Y., Landry, S.J., 2003. The major peanut allergen, Ara h 2, functions as a trypsin inhibitor, and roasting enhances this function. *J. Allergy Clin. Immunol.* 112, 190–195.
- McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., Read, R.J., 2005. Likelihood-enhanced fast translation functions. *Acta Cryst. D* 61, 458–464.
- Merritt, E.A., Murphy, M.E., 1994. Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Cryst. D* 50, 869–873.
- Muntz, K., 1998. Deposition of storage proteins. *Plant Mol. Biol.* 38, 77–99.
- Nowak-Wegrzyn, A., 2007. Food allergy to proteins. *Nestle Nutr. Workshop Ser. Pediatr. Program* 59, 31–36, 17–31; discussion.
- Otwinowski, Z., Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. In: Carter, C.W., Sweet, R.M. (Eds.), *Methods in Enzymology*, 276. Academic Press, New York, pp. 307–326.
- Ozol, D., Mete, E., 2008. Asthma and food allergy. *Curr. Opin. Pulm. Med.* 14, 9–12.
- Pons, L., Palmer, K., Burks, W., 2005. Towards immunotherapy for peanut allergy. *Curr. Opin. Allergy Clin. Immunol.* 5, 558–562.
- Potterton, E., Briggs, P., Turkenburg, M., Dodson, E., 2003. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* 59, 1131–1137.
- 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.
- Sampson, H.A., 2004. Update on food allergy. *J. Allergy Clin. Immunol.* 113, 820, 805–819; quiz.
- Sampson, H.A., Mendelson, L., Rosen, J.P., 1992. Fatal and near-fatal anaphylactic reactions to food in children and adolescents. *N. Engl. J. Med.* 327, 380–384.
- Sayle, R.A., Milner-White, E.J., 1995. RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* 20, 374.
- Shreffler, W.G., Beyer, K., Chu, T.H., Burks, A.W., Sampson, H.A., 2004. Microarray immunoassay: association of clinical history, in vitro IgE function, and heterogeneity of allergenic peanut epitopes. *J. Allergy Clin. Immunol.* 113, 776–782.
- 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.
- Storoni, L.C., McCoy, A.J., Read, R.J., 2004. Likelihood-enhanced fast rotation functions. *Acta Cryst. D* 60, 432–438.
- Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., Murshudov, G.N., 2004. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2184–2195.
- Viquez, O.M., Konan, K.N., Dodo, H.W., 2004. Genomic organization of peanut allergen gene, Ara h 3. *Mol. Immunol.* 41, 1235–1240.
- Wang, F., Robotham, J.M., Teuber, S.S., 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.
- Wang, J., Sampson, H.A., 2007. Food anaphylaxis. *Clin. Exp. Allergy* 37, 651–660.
- Wensing, M., Knulst, A.C., Piersma, S., O’Kane, F., Knol, E.F., Koppelman, S.J., 2003. Patients with anaphylaxis to pea can have peanut allergy caused by cross-reactive IgE to vicilin (Ara h 1). *J. Allergy Clin. Immunol.* 111, 420–424.
- Yan, Y.-S., Lin, X.-D., Zhang, Y.-S., Wang, L., Wu, K., Huang, S.-Z., 2005. Isolation of peanut genes encoding arachins and conglutins by expressed sequence tags. *Plant Sci.* 169, 439–445.
- Yunginger, J.W., Sweeney, K.G., Sturmer, W.Q., Giannandrea, L.A., Teigland, J.D., Bray, M., Benson, P.A., York, J.A., Biedrzycki, L., Squillace, D.L., et al., 1988. Fatal food-induced anaphylaxis. *JAMA* 260, 1450–1452.