



Contents lists available at ScienceDirect

Molecular Immunology

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



Crystal structure of the major peanut allergen Ara h 1

Cerrone Cabanos^a, Hiroyuki Urabe^a, Mary Rose Tandang-Silvas^a, Shigeru Utsumi^a,
Bunzo Mikami^b, Nobuyuki Maruyama^{a,*}

^a Laboratory of Food Quality Design and Development, Graduate School of Agriculture, Kyoto University Uji, Kyoto, Japan

^b Laboratory of Applied Structural Biology, Graduate School of Agriculture, Kyoto University Uji, Kyoto, Japan

ARTICLE INFO

Article history:

Received 27 June 2011

Received in revised form 30 July 2011

Accepted 6 August 2011

Available online xxx

Keywords:

Peanut

Food allergen

IgE epitope

X-ray crystallography

7S globulin

Electrostatic potential

Thermal stability

ABSTRACT

Ara h 1, a 7S globulin, is one of the three major peanut allergens. We previously reported the crystallization of the core region of recombinant Ara h 1. Here, we present the crystal structure of the Ara h 1 core at a resolution of 2.43 Å. We also assayed the Ara h 1 core thermal stability and compared its final structure against other 7S globulins. The Ara h 1 core has a thermal denaturation temperature of 88.3 °C and a structure that is very similar to other 7S globulins. Previously identified linear IgE epitopes were also mapped on the three-dimensional structure. Most linear epitopes were found in the extended loop domains and the coils between the N- and C-terminal modules, while others were found in the less accessible β-sheets of the C-terminal core β-barrel domain of each monomer. Most of these epitopes become either slightly or significantly buried upon trimer formation, implying that allergen digestion in the gut is required for these epitopes to be accessible to immunoglobulins. Our findings also suggest that both intact and partially degraded allergens should be employed in future diagnostic and immunotherapeutic strategies.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The peanut seed and its food derivatives belong to the “big eight” group of foods that account for the majority of food allergies worldwide, along with milk, eggs, fish, crustaceans, wheat, tree nuts, and soybean products. Peanut allergy has a worldwide prevalence of 0.5–2%, and this prevalence appears to be increasing (Bindslev-Jensen et al., 2002; Emmett et al., 1999; Grundy et al., 2002; Osterballe et al., 2005; Sicherer et al., 1999, 2003). The symptoms of peanut allergy range from mild oral allergy syndrome (OAS) to anaphylactic reactions, and even death. As yet, there is no cure for peanut allergy. Therapy focuses primarily on peanut avoidance, early recognition of symptoms brought on by accidental ingestion, and pharmacologic treatment of adverse reactions (Bock et al., 2001; Roberts, 2007; Skolnick et al., 2001). To date, there are 13 listed allergens that account for peanut allergy, including 11 allergens named Ara h 1 through 11, and two recently identified allergens, agglutinin and 18-kDa oleosin (Mari et al., 2006). The names for allergens Ara h 1–11 are officially recognized by the Allergen Nomenclature Sub-Committee of the International Union of Immunological Societies (IUIS).

The peanut seed contains both storage and non-storage proteins. Allergens Ara h 1–4, Ara h 6, and Ara h 7 are seed storage

proteins. The major peanut allergens are Ara h 1–3. Ara h 1 is a 7S vicilin-like globulin also known as conarachin (de Jong et al., 1998; Kleber-Janke et al., 1999; Wichers et al., 2004). Ara h 2 is a 2S albumin, also known as conglutin, and functions as a trypsin inhibitor (Maleki et al., 2003). Ara h 3 is an 11S globulin known as arachin, and also functions as a trypsin inhibitor (Dodo et al., 2004). The sequence identity between Ara h 2 and Ara h 6 is 59%, while that between Ara h 2 and Ara h 7 is 35% (Lehmann et al., 2006). Ara h 2, h 6, and h 7 are regarded as isoforms, as are Ara h 3 and Ara h 4 (Boldt et al., 2005; Koppelman et al., 2003). Non-storage proteins contained within peanut seeds include Ara h 5, Ara h 8–11, agglutinin, and 18-kDa oleosin. Ara h 5 is an actin-binding protein that is also known as profilin. Ara h 8 is a pathogenesis-related protein, also called PR-10. Ara h 9 is a lipid transfer protein, while Ara h 10 and Ara h 11 are oleosins, or oil body proteins. Agglutinin is a lectin involved in cellular recognition (Mari et al., 2006).

Since the allergenicity of a particular protein largely depends on its overall structure, the development of advanced methods for the diagnosis and treatment of peanut allergy requires a better understanding of the three-dimensional structure of all the allergens involved. However, to date, solved structures are available for only Ara h 2, Ara h 3, Ara h 6, and agglutinin. The objective of this study was to elucidate the structure of the remaining major peanut allergen, Ara h 1. This 65-kDa protein has a high sequence similarity with other plant vicilins, which implies that there is a high probability of cross-reactivity. Ara h 1 comprises 12–16% of the total peanut protein (de Jong et al., 1998), and affects 35–95% of peanut-allergic

* Corresponding author. Tel.: +81 774 39 5790, fax: +81 774 38 3761.

E-mail address: marunobu@kais.kyoto-u.ac.jp (N. Maruyama).

patients in different populations (Mari et al., 2006). Size exclusion chromatography has revealed that Ara h 1 exists as a stable trimer during recombinant expression (Pomes et al., 2003), but exists as an oligomer upon purification of the native protein from seeds (van Boxel et al., 2006).

Initially, we attempted to crystallize the mature Ara h 1 protein; however, due to the difficulty in obtaining crystals with good diffraction quality, the N-terminal extension and C-terminal flexible regions that can negatively affect diffraction were removed. The study thus focused on the Ara h 1 core region instead. Previously, we reported the cloning, expression in *Escherichia coli*, purification, and crystallization of the Ara h 1 core region (Cabanos et al., 2010). Here, we present the solved three-dimensional structure of the Ara h 1 core region and mapping of the known linear epitopes on its molecular surface.

2. Materials and methods

2.1. Cloning, expression, and purification of the Ara h 1 core region

Details regarding the cloning, expression, and purification of the Ara h 1 core region have been described previously (Cabanos et al., 2010). Briefly, the Ara h 1 core region cDNA was ligated to pET21d (Novagen, CA) and transformed into *E. coli* BL21(DE3) (Novagen). The bacteria were cultured in Luria-Bertani (LB) medium containing 0.5 M NaCl and 50 µg/ml carbenicillin and expression of the recombinant protein was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) (1 mM final concentration) when the culture reached an OD_{600 nm} = 0.60. After induction, cultures were incubated for an additional 48 h at 20 °C. Cells were then harvested, resuspended, and sonicated in buffer A [35 mM potassium phosphate (pH 7.4), 0.4 M NaCl, 1 mM EDTA, 0.1 mM (*p*-amidinophenyl)-methanesulfonyl fluoride (*p*-APMSF), 1.2 µM leupeptin, 0.2 µM pepstatin A, 0.02% (w/v) NaN₃]. The resulting soluble fraction was subjected to a three-step purification involving ammonium sulfate [(NH₄)₂SO₄] fractionation, ion-exchange chromatography using a Q-Sepharose HR 26/10 column (GE Healthcare Life Sciences, NJ), and gel-filtration chromatography using a HiLoad 26/60 Superdex 200 prep-grade column (GE Healthcare Life Sciences). The purity of Ara h 1 core (47.4 kDa) fractions was estimated using Laemmli (1970) 11% SDS-PAGE gels.

2.2. Protein measurement

Protein concentrations were determined using a Protein Assay Rapid Kit (Wako, Japan), with bovine serum albumin (BSA) as the standard.

2.3. Analysis of self-assembly into trimers

Self-assembly of the recombinant Ara h 1 core was analyzed using a Hi-Prep 16/60 Sephacryl S-300 HR column (GE Healthcare Life Sciences). The flow rate was set at 0.5 ml/min using the same buffer described above (buffer A).

2.4. Differential scanning calorimetry

Differential scanning calorimetry (DSC) experiments were carried out on a Microcal MC-2 Ultrasensitive Microcalorimeter (Microcal Inc., NJ). The software package Origin, supplied by Microcal, was used for data analysis and plotting. The DSC scan rate was set at 1 °C/min. All DSC experiments were performed at a protein concentration of 0.5 mg/ml using buffer A.

2.5. Crystallization, structure solution, and refinement

Purified recombinant Ara h 1 core in buffer A was concentrated to 10 mg/ml. Screening for crystallization conditions was carried out in 24-well Linbro plates using the hanging-drop vapor-diffusion technique against a 1 ml well solution. Diffraction-quality crystals were obtained in a reservoir containing 0.1 M sodium citrate (pH 5.6), 0.1 M NaCl, and 15% PEG400 at 20 °C. X-ray diffraction images of the crystal were collected using a JUPITER 210 CCD detector (Rigaku, Japan) with synchrotron radiation at a wavelength of 1.0 Å at the BL-38B1 station of SPring-8 (Hyogo, Japan). Diffraction data were indexed, integrated, and scaled using the HKL2000 program. The initial structure was determined using the molecular-replacement technique with the MOLREP (Vagin and Teplyakov, 2010) program from the CCP4 suite [Collaborative Computational Project, Number 4, (1994)] and using the rigid-body refinement option with the β-subunit of soybean β-conglycinin (PDB: 1IPK) as the template. Final model building and refinement were performed using PHENIX (Adams et al., 2010) and Coot (Emsley and Cowtan, 2004), respectively.

2.6. Structural analysis

The refined model was validated with a Ramachandran plot calculated using PROCHECK (Vaguine et al., 1999). Structural alignment was carried out using the Secondary Structure Matching (SSM) Superpose program (Krissinel and Henrick, 2004) under Coot. Secondary structures were determined using DSSP and validated using Coot's SSM Superpose. The surface accessibility of residues was calculated using both DSSP (Kabsch and Sander, 1983) and AREAIMOL (Lee and Richards, 1971). The molecular surfaces were calculated and displayed using PyMOL. Electrostatic potentials were calculated and displayed using GRASP2 (Petrey and Honig, 2003).

3. Results

3.1. Self-assembly of the Ara h 1 core

The purity of the recombinant Ara h 1 core was estimated to be above 95%, as described previously (Cabanos et al., 2010). Gel filtration chromatography experiments confirmed that recombinant Ara h 1 core exists as a trimer (~145 kDa; data not shown), similar to native Ara h 1.

3.2. Ara h 1 core model quality

The Ara h 1 core region amino acid sequence compared to other 7S globulins with known crystal structures is shown in Fig. 1. The sequence of Ara h 1 is 53.0% identical to that of soybean β-conglycinin β (PDB: 1IPK), 52.0% identical to mung bean 8S globulin (PDB: 2CV6), 50.1% identical to both soybean β-conglycinin α' (PDB: 1UIK) and adzuki 7S1 (PDB: 2EA7), and 49.1% identical to adzuki 7S3 (PDB: 2EAA).

The crystal structure of the Ara h 1 core region was determined at a final resolution of 2.43 Å. The data collection and refinement statistics are presented in Table 1. A total of 1732 (90.1%) of the amino acids were within the core favored region and 1917 (99.7%) of the amino acids, excluding glycine (130), proline (96) and terminal residues (55), were within the allowed region of the Ramachandran plot of the refined model as calculated by PROCHECK, while 0.3% (6) of the amino acids were found in the outlier region. The outlier residue His412, found in six subunits, is clearly defined in the density map and exists in the middle of the γ-turns. The *R* and *R*_{free} values from the resulting model were 20.2% and 24.4%, respectively. The asymmetric

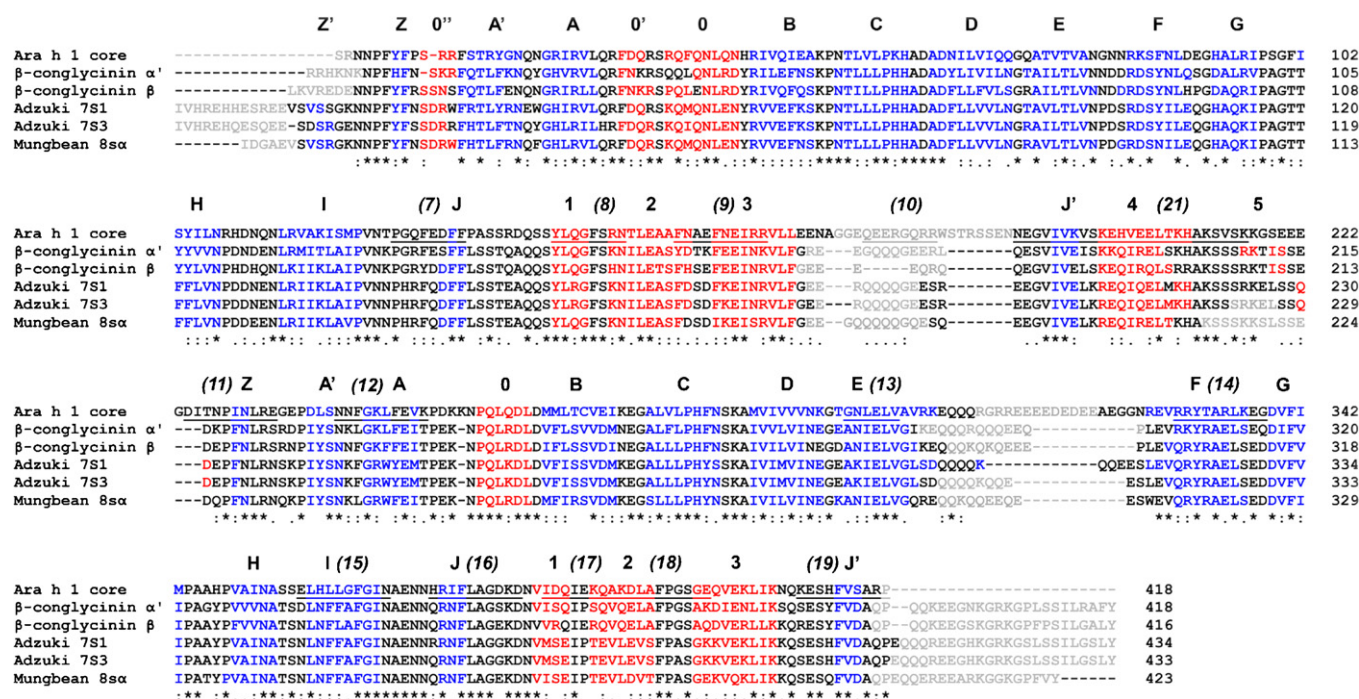


Fig. 1. Structure-based sequence alignment of the Ara h 1 core with other solved structures of 7S globulins, generated using the Secondary Structure Matching (SSM) Superspace program (Krissinel and Henrick, 2004) under Coot. Residues not found in the final model are shown in gray. Blue and red represent β -strands and α -helices, respectively. The symbols (-) and (*) represent gaps and conserved residues, respectively, while (.) and (:) indicate the relative degree of homology among residues. Letters and numbers at the top of the alignment are the annotation of the typical secondary-structure elements of 7S proteins, adopted from Maruyama et al. (2001). Numbers in parentheses refer to the name/number of previously identified linear epitopes (underlined) of Ara h 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

unit of the crystal contained two trimers, hence, six monomers or chains, which were designated as A–F, and also contained a total of 337 water molecules. A total of 373 residues were present in monomers A and D, 365 residues in monomers B and E, 370

residues in monomer C, and 364 in monomer F. In all of the Ara h 1 core monomers, the N-termini (residue 1 of monomer A and residues 1–2 of monomers B–F) and C-termini (residue 417 of monomers A–D, and F and residues 417–418 of monomer E) were not included in the final models. Furthermore, three major regions were not included in most of the monomers due to their diffused 2|Fo| – |Fc| maps, leading to difficulty in tracing the correct residues in these regions. These regions also generally coincide with the missing regions in the crystal structures of other 7S globulins (Fig. 1). Most of the residues in the missing regions are hydrophilic and probably located at the surface, indicating that these are highly disordered regions. The first missing region corresponds to residues 172–190 of monomer A, residues 173–191 of monomer C, residues 172–191 of monomers B and D, residues 172–192 of monomer E, and residues 172–193 of monomer F. The second missing region corresponds to residues 215–218 of monomer A, residues 215–221 of monomer B, residues 216–220 of monomer C, residues 219–222 of monomer D, residues 214–217 of monomer E, and residues 214–220 of monomer F. The third missing region corresponds to residues 306–325, 304–326, 304–324, 305–322, 303–326, and 304–325 of monomers A–F, respectively. Each monomer also includes an oxidized cysteine at position 266 (Fig. 2C). In this study, monomer D was used in all figures as it had the most visible electron density map and fewest missing regions.

3.3. Overall structure

The overall trimeric structure of the Ara h 1 core region is shown as a ribbon model in Fig. 2A. The core region has approximate dimensions of $95 \text{ Å} \times 95 \text{ Å} \times 40 \text{ Å}$, and a monomeric model is shown in Fig. 2B. The monomer can be divided into two very similar modules, the N- and C-terminal modules, separated by a pseudo-dyad axis. Each module contains the conserved β -barrel core (jelly roll-like) domain and an extended loop domain con-

Table 1

Summary of data collection and refinement statistics.

Data collection	
Wavelength (Å)	1.0
Temperature (K)	100
Space group	C2
a, b, c (Å)	156.49, 89.01, 158.97, 90.00, 107.15, 90.00
Resolution limits (Å)	2.43 (2.46–2.43)
Number of observed reflections	757113 (70174.2)
Number of unique reflections	122061 (12099)
Completeness (%)	100 (99.9)
Mean I/I(σ)	14.0 (5.45)
R _{merge} (%)	7.9 (43.6)
Structure refinement statistics	
Resolution range (Å)	15.0–2.43 (2.46–2.43)
Number of reflections	76734 (2554)
Working set	72875 (3616)
Test set(5.03%)	3859 (130)
Number of atoms	
Protein	18099
Water	596
R/R _{free} (%)	20.2/24.4 (26.7/34.4)
Rmsd from ideal geometry	
Bond lengths (Å)	0.008
Bond angles (°)	1.072
Ramachandran plot (%)	
Number of residues in favored region	1732 (90.1% of non-Gly and non-Pro)
Number of residues in allowed region	185 (9.6% of non-Gly and non-Pro)
Number of residues in outlier region	6 (0.3% of non-Gly and non-Pro)

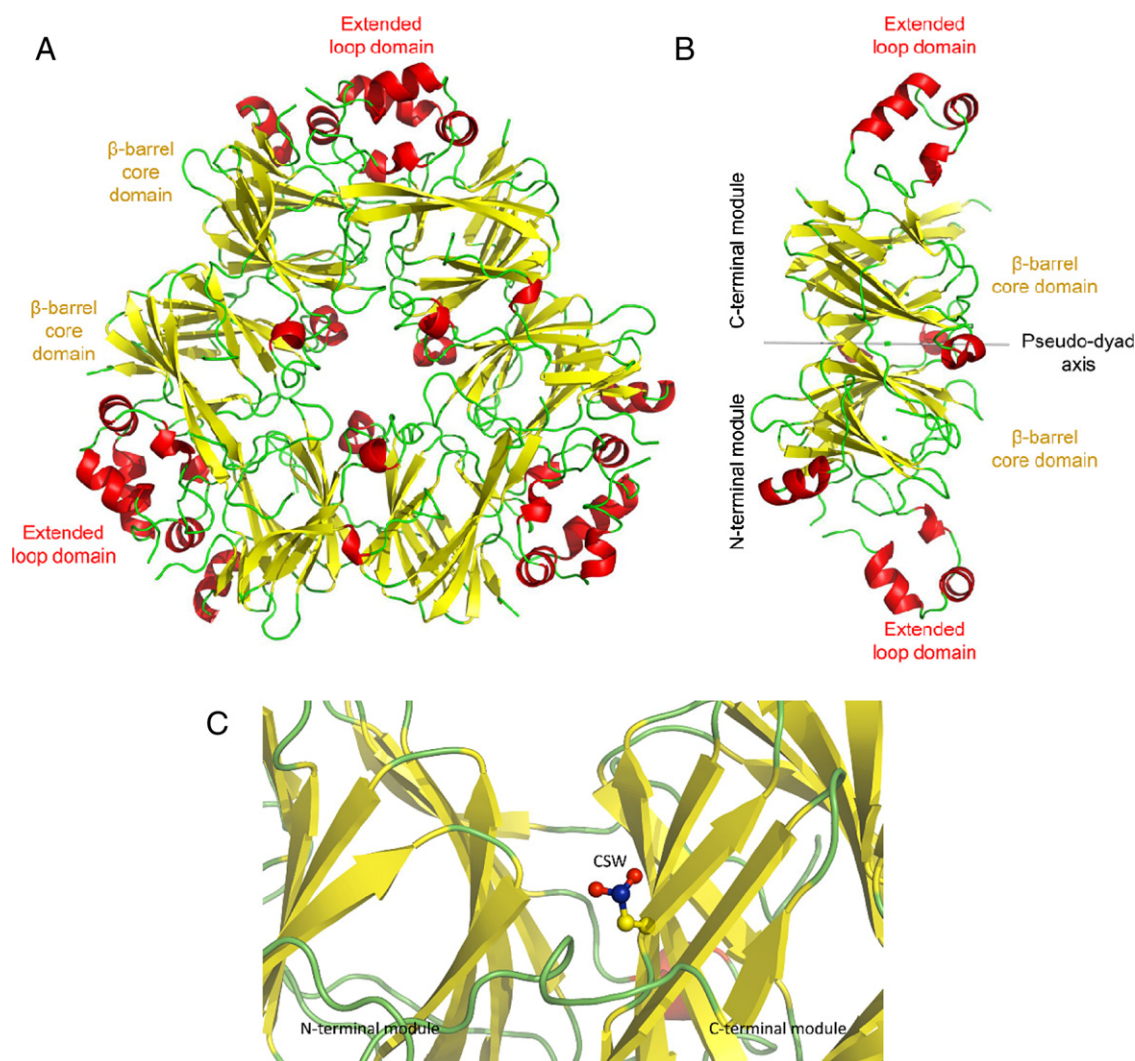


Fig. 2. Overall structure of the Ara h 1 core. (A) Trimer structure. (B) Monomer structure. (C) Oxidized cysteine (CSW) at position 266 of each monomer. Strands of β -sheets and stretches of α -helices are shown in yellow and red, respectively. The gray line represents the pseudo-dyad axis in between the N- and C-terminal modules. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

taining several helices. Each β -barrel core domain consists of two β -sheets made up of A'ABIDG and J'JCHEF segments (Fig. 1). These secondary-structure elements were named according to the strands and helices in soybean β -conglycinin β (Maruyama et al., 2001).

3.4. Structural comparison

The C α backbone superimposition of Ara h 1 with other known structures of 7S globulins is shown in Fig. 3. Although there are minor differences, the Ara h 1 core backbone structure is very similar to those of β -conglycinin α' , β -conglycinin β , adzuki 7S1 and 7S3, and mung bean 8S α . The Ara h 1 core has root mean square deviations (RMSD) of 1.15 Å compared to the 358 C α atoms of β -conglycinin α' , 1.01 Å for the 359 C α atoms of β -conglycinin β , 1.04 Å for the 367 C α atoms of adzuki 7S1, 1.00 Å for the 361 C α atoms of adzuki 7S3, and 1.04 Å for the 352 C α atoms of mung bean 8S α . Minor differences among 7S globulins (Figs. 1 and 3) are found at the N-termini, N-terminal extended loop domains (between segments 0 and 0'), variable region in between segments 3 and J' of N-terminal modules, random coils in between N- and C-terminal modules (segments 4 and Z), between segments A and 0' of C-

terminal modules, the variable region in between segments E and F of C-terminal modules, and at the C-termini.

3.5. Ara h 1 thermal stability and surface character

The thermal stability of the Ara h 1 core in terms of its midpoint denaturation temperature (88.3 °C) compared to other globulins (Fukuda et al., 2008) is shown in Table 2. As shown, the Ara h 1 core T_m falls within the range observed for other 7S globulins (77.5–90.8 °C) and is above the average (87.5 °C), indicating that it has a fairly high thermal stability. The order of T_m values was adzuki 7S3 $\geq$ adzuki 7S1 $>$ β -conglycinin β $>$ Ara h 1 core $>$ β -conglycinin α' $>$ mung bean 8S α .

The distributions of aromatic (Phe, Tyr, and Trp) and aliphatic (Ala, Gly, Ile, Leu, Met, Pro, and Val) hydrophobic residues on the surface of the Ara h 1 core and its surface electrostatic potential in the trimeric and monomeric forms are shown in Figs. 4 and 5, respectively. Overall, the surface distribution of both aromatic and aliphatic hydrophobic residues (Figs. 4A and 5A) in the Ara h 1 core is very similar to that of the other 7S proteins, as previously shown by Fukuda et al. (2008). However, in terms of surface electrostatic potential, Ara h 1 seems to possess the largest distribution

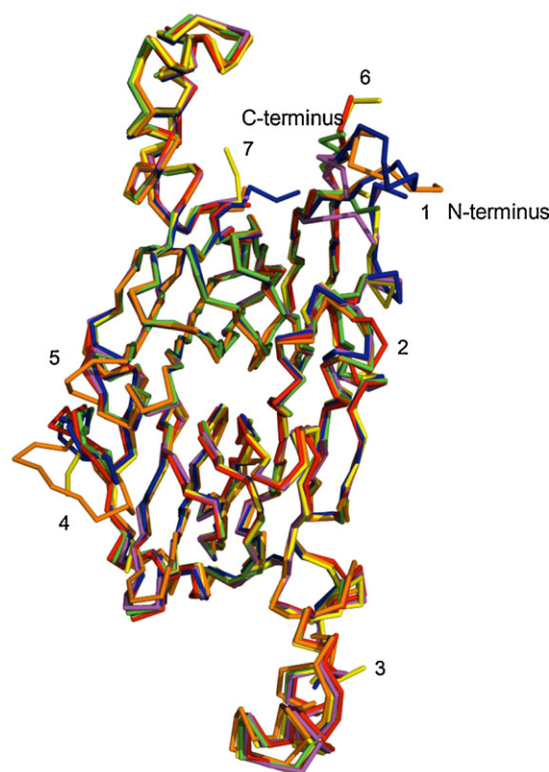


Fig. 3. α backbone superimposition of the Ara h 1 core with other 7S globulins. Ara h 1 core, β -conglycinin α' , β -conglycinin β , adzuki 7S1, adzuki 7S3 and mungbean 8S α are shown in orange, red, green, blue, yellow, and magenta, respectively. 1: N-terminus; 2: N-terminal extended loop domain (between segments 0 and 0'); 3: variable region in between segments 3 and J' of the N-terminal module; 4: random coil in between the N- and C-terminal modules (segments 4 and Z); 5: random coil in between A and O' of the C-terminal module; 6: variable region in between segments E and F of the C-terminal module; 7: C-terminus.

of electropositive surface at the center of one of the sides (Fig. 4B) of its trimeric form, which is then followed by β -conglycinin α' and β -conglycinin β (data shown in Fukuda et al., 2008). In all three proteins, the center and periphery of one surface have well-defined positive and negative potentials, respectively, and the center and periphery of the other surface have negative and positive potentials, respectively. The other 7S globulins appear to have a less concentrated distribution of either electropositive or electronegative potentials on their surfaces.

3.6. Map of the linear epitopes on the Ara h 1 core surface

To date, 24 Ara h 1 linear epitopes have been identified (Burks et al., 1997; Shin et al., 1998; Shreffler et al., 2004). Based on the recently reported correct N-terminal sequence of native Ara h 1 (Wichers et al., 2004), the first three previously identified epitopes are now considered irrelevant because they are located in the signal peptide region. Relative to mature Ara h 1, there are 21 epitopes (numbered 1–21) that span the entire sequence. The data listed in Table 3 show the previously reported linear epitopes (see also Fig. 1) found in the Ara h 1 core region, as well as their corresponding solvent accessible surface areas (these are epitopes 7–21, excepting epitope 10, which is located in the disordered region, and epitope 20, which is in the omitted C-terminal flexible region). The localization of these epitopes on the molecular surface of the Ara h 1 trimeric and monomeric forms is shown in Figs. 4C and 5C, respectively. Upon trimer formation (Fig. 4C; Table 3), all epitopes except epitope 12 become partially (<50%) to significantly ($\geq 50\%$) buried. Significantly buried epitopes include

Table 2

Thermal stability of the Ara h 1 core and other 7S globulins.

7S globulins	T_m values ($^{\circ}\text{C}$)
Ara h 1 core	88.3
β -Conglycinin α'	83.3 [#]
β -Conglycinin β	90.8 [#]
Adzuki 7S1	92.4 [#]
Adzuki 7S3	92.5 [#]
Mung bean 8S α	77.5 [#]

[#] These data are cited from Fukuda et al. (2008), Itoh et al. (2006) and Maruyama et al. (1998).

numbers 7 (63.0% burial), 13 (56.6%), 17 (64.6%), and 18 (61.2%), while epitopes 8 (46.5% burial), 9 (28.3%), 11 (17.8%), 14 (7.0%), 15 (8.9%), 16 (43.7%), 19 (6.6%), and 21 (10.2%) are partially covered but still solvent accessible. Epitopes 7, 13, 17, and 18 contain a high number of hydrophobic residues (Fig. 4A and C; Table 3), and closer inspection revealed that these epitopes are also located in regions where actual monomer–monomer interactions occur during trimer formation (Figs. 4C and 5C).

Epitope 15 (shown in orange in Figs. 4C, 5C and D) is already significantly buried inside the C-terminal β -barrel domain (Table 3; Fig. 5C and D) in the monomeric form of Ara h 1, with 80% of its residues buried and only 2 of its 10 residues accessible to solvent (underlined residues in Table 3). The accessible surface area (ASA) of epitope 15 is one of the lowest among the Ara h 1 epitopes, at only 194.2 $\text{\AA}^2$. Epitope 21 has the largest ASA (1703.6 $\text{\AA}^2$), while the ASA of epitope 13 (179.8 $\text{\AA}^2$) is the lowest. Epitope 13 is located in one of the outer β -sheets of the C-terminal β -barrel core domain, while epitope 21 is located in the β -sheet-coil-helix segment on the periphery of the N-terminal β -barrel core domain. The linear epitopes (nos. 7–12, 14–19) are reported to contain residues that are critical for IgE binding (Burks et al., 1997), and although most of these critical residues are generally accessible to solvent in the monomer form, some are completely inaccessible.

Epitopes 7–9 and 17–19 are located in the conserved N- and C-extended loop regions, respectively. Some epitopes are located in the β -sheets on the periphery of the C-terminal β -barrel core (nos. 12, 14, and 16), while epitope 11 is located in the coils in between the two β -barrel core domains (Fig. 5D). Except for epitopes 7, 13, 15, 17, and 18, most of the epitopes are located in relatively hydrophilic regions on the surface of the Ara h 1 core (Figs. 4A and 5A). Moreover, epitopes 9, 14, 19, and 21 are located in relatively neutral to electropositive regions of the Ara h 1 surface, while the remaining epitopes are located in neutral to electronegative regions.

4. Discussion

The peanut 7S globulin Ara h 1 was the first peanut allergen identified (Burks et al., 1991) and the first of the three immunodominant peanut allergens described. Elucidating the structure of Ara h 1 is critical for understanding the allergenicity and possible cross-reactivity of Ara h 1 with other 7S globulins. The mature Ara h 1 protein has three major structural characteristics in common with other 7S globulins (Fukuda et al., 2008; Itoh et al., 2006; Maruyama et al., 2001, 2004): the N-terminal extension region (residues 1–86, based on the correct N-terminal sequence of native Ara h 1 reported by Wichers et al. (2004)), the core region (residues 86–504), and the C-terminal flexible region (residues 504–536). The sequences of both terminal regions are variable, while that of the core region is highly conserved among 7S globulins.

Initially, our group attempted to recombinantly produce and crystallize the Ara h 1 mature protein. However, after several months of screening trials, no diffraction-quality crystals were obtained. After removing the extension and flexible regions of Ara

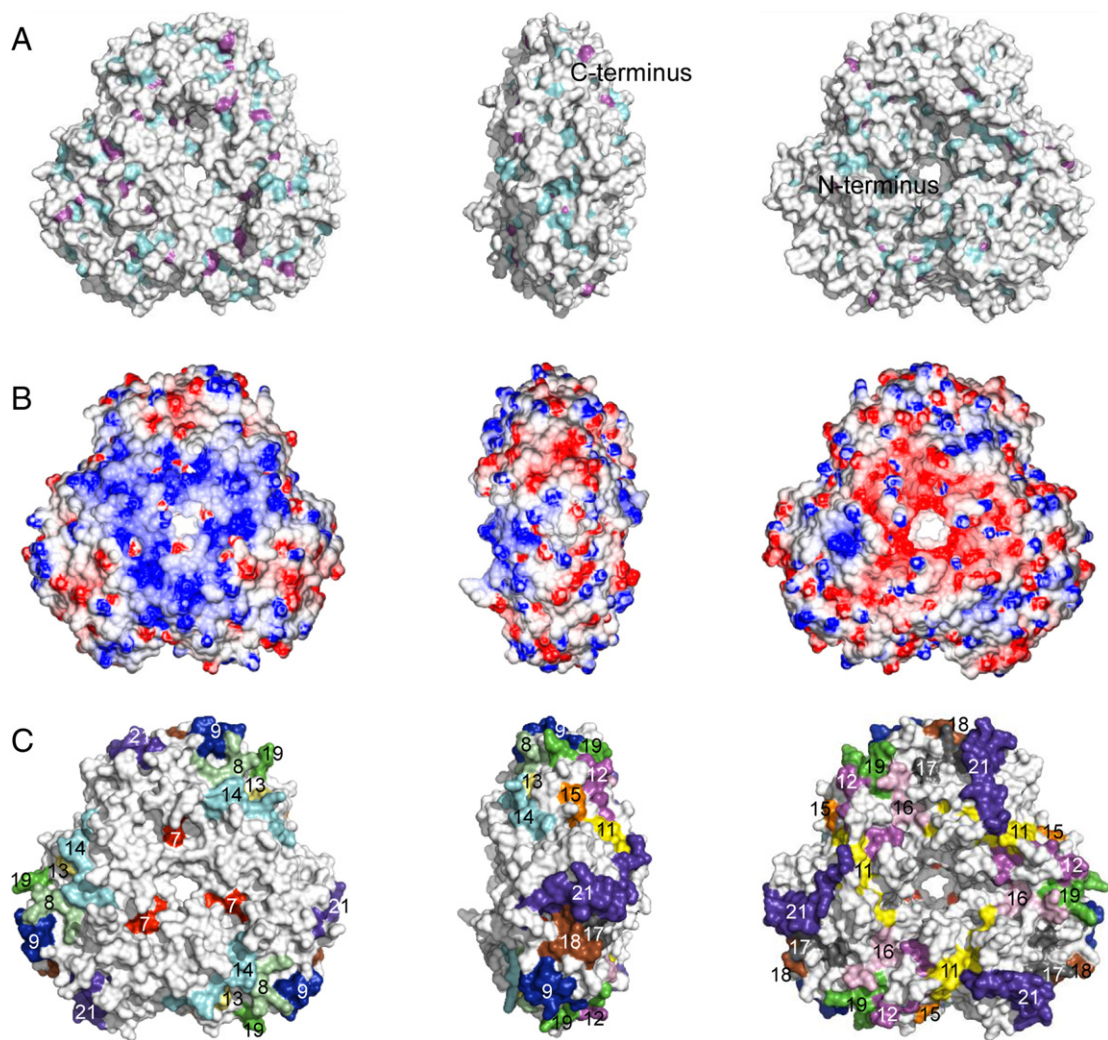


Fig. 4. Surface character and mapping of previously identified linear IgE epitopes on the molecular surface of the Ara h 1 core trimer. (A) Distribution of aromatic (Phe, Tyr, and Trp) and aliphatic (Ala, Gly, Ile, Leu, Met, Pro, and Val) hydrophobic residues on the molecular surface are shown in pink and cyan, respectively. (B) Ara h 1 core trimer surface electrostatic potential as calculated using GRASP2 and rendered from the -10 kT (red; electronegative) to $+10\text{ kT}$ (blue, electropositive) range, where k is the Boltzmann constant and T is the absolute temperature (K). (C) Map of previously identified linear IgE epitopes. Epitopes are shown in different colors: no. 7: red; 8: pale green; 9: blue; 11: yellow; 12: magenta; 13: pale yellow; 14: cyan; 15: orange; 16: light pink; 17: dark gray; 18: brown; 19: green; and 21: purple. Structures are represented as white molecular surface models using Pymol. The center and rightmost diagrams represent views after 90° and 180° rotation around the vertical axis, respectively, of the leftmost diagram. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Known linear IgE epitopes in the Ara h 1 core region and their corresponding solvent accessible surface areas (ASA) in the monomeric and trimeric forms.

Epitope no. ^a	Epitope sequence	Residue no. based on Ara h 1 core sequence	Trimer ASA (Å ²)	Monomer ASA (Å ²)	% buried area in trimeric form
7 ^a	PGQFEDFF	126–133	136.5	369.1	63.0
8 ^a	YLQGFSRN	143–150	360.9	675.2	46.5
9 ^a	FNAEFNEIRR	156–165	600.9	837.5	28.3
10 ^a	QEERGQRR	176–183 (disordered)			
11 ^a	DITNPINLRE	224–233	309.5	376.4	17.8
12 ^a	NNFGKLFVVK	240–249	458.4	458.4	0.0
13 ^a	GNLELV	294–299	78.0	179.8	56.6
14 ^a	RRYTARLKEG	329–338	609.9	655.6	7.0
15 ^a	ELHLLGFGIN	356–365	177.0	194.2	8.9
16 ^a	HRIFLAGDKD	370–379	314.3	558.5	43.7
17 ^a	IDQIEKQAKD	382–391	264.7	746.9	64.6
18 ^a	KDLAFPGSGE	390–399	256.7	661.7	61.2
19 ^a	KESHFVSARP	409–418	566.1	606.2	6.6
21 ^b	NEGVIVKVSKEHVEELTKHAKSVSK	192–216	1529.8	1703.6	10.2

Bold face letters are the critical residues for IgE binding (Shin et al., 1998).
Underlined residues are surface-exposed or solvent accessible in monomer determined using DSSP (Kabsch and Sander, 1983) and AREAIMOL (Lee and Richards, 1971).
^a Determined by Burks et al. (1997).
^b Determined by Shreffler et al. (2004).
^{*} Numbered according to van Bostel et al. (2008).

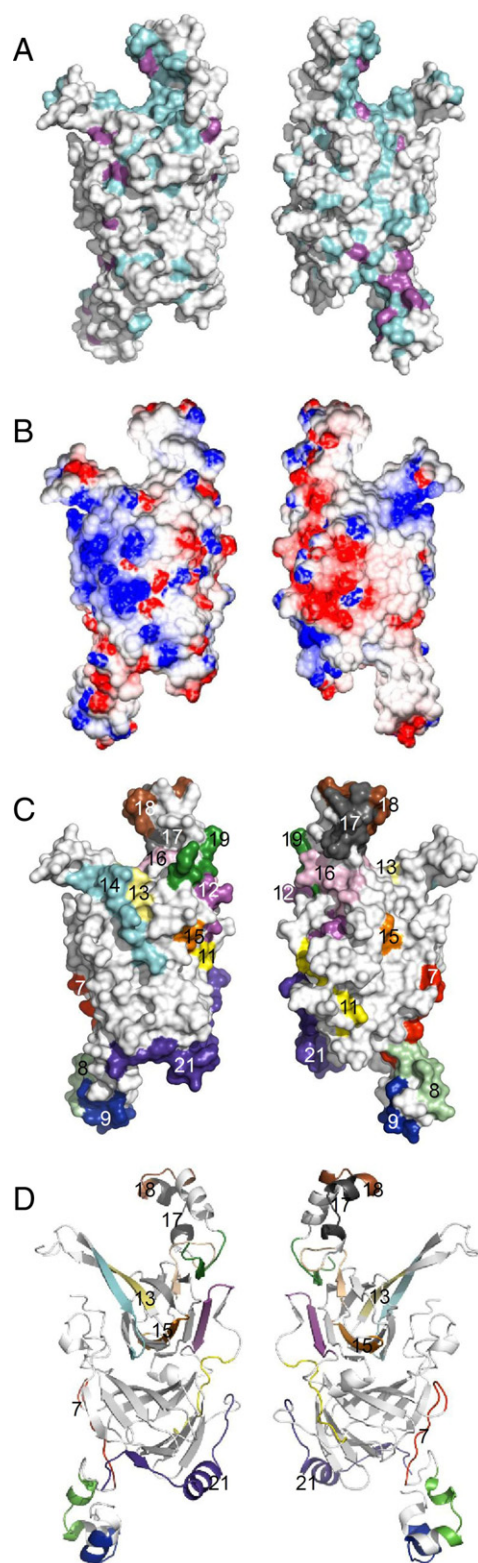


Fig. 5. Surface character and mapping of previously identified linear IgE epitopes on the Ara h 1 core monomer, and their secondary structural features. (A) Distribution of aromatic (Phe, Tyr, and Trp) and aliphatic (Ala, Gly, Ile, Leu, Met, Pro, and Val) hydrophobic residues on the Ara h 1 core monomer molecular surface. (B) The Ara h 1 core monomer surface electrostatic potential. (C) Map of previously identified linear IgE epitopes. (D) Secondary structural features of the epitopes. The left and right diagrams represent the 0° and 180° rotations, respectively, around the vertical axis. The hydrophobicity, surface electrostatic potential, map of epitopes on surface, and secondary structure views were rendered in the same manner as those of Fig. 4.

h 1, the crystal quality improved, and we recently reported the expression and crystallization of the Ara h 1 core region (Cabanos et al., 2010). In this study, we present the refined crystal structure of the Ara h 1 core region at a resolution of 2.43 Å (Table 1). The Ara h 1 allergen was found to exist in trimeric form (Fig. 2; gel-filtration data not shown) and its topology and basic structure are very similar to other known structures of 7S globulins, namely, soybean β -conglycinin α' and β , adzuki 7S1 and 7S3, and mung bean 8S α (Figs. 1 and 3). Ara h 1 was also found to have a surface distribution of hydrophobic residues very similar to that of the above-mentioned globulins (Figs. 4A and 5A). Obviously, these similarities indicate that there is a high possibility of cross-reactivity between 7S globulins.

Although there are many similarities among these proteins, some differences were also found, particularly in regard to thermal stability and electrostatic potential. In this study, we also determined the thermal denaturation temperature (T_m) of the Ara h 1 core (Table 2) and compared it to the reported T_m values for other 7S globulins (Fukuda et al., 2008; Maruyama et al., 1998). The T_m of the Ara h 1 core is 88.3 °C, which falls within the range reported for other 7S globulins (77.5–90.8 °C). This value is also above the average T_m (87.5 °C) of the other 7S proteins, indicating that Ara h 1 has a relatively high thermal stability. A high thermal stability can also be taken to indicate high overall protein stability. It is known that a protein's stability can greatly influence its resistance to digestion. Since resistance to digestion is particularly relevant to a protein's role as an allergen, a high stability implies high allergenic potency (Astwood et al., 1996; Mills et al., 2002; Sen et al., 2002). The order of 7S globulin thermal stability can be summarized as follows: adzuki 7S3 $\geq$ adzuki 7S1 $>$ β -conglycinin β $>$ Ara h 1 core $>$ β -conglycinin α' $>$ mung bean 8S α . This could also mean that Ara h 1 may have a higher allergenicity than the other 7S globulins with lower T_m values.

In terms of surface electrostatic potential profile, among the five 7S globulins Ara h 1 (Figs. 4B and 5B) has the largest distribution of electropositive surfaces at the center of one of its sides, which is then followed by that of β -conglycinin α' and β -conglycinin β . Variation in the electrostatic potential of a protein can influence the affinity and specificity of IgEs directed against these proteins. Moreover, since most of the IgE epitopes tend to be located within the electropositive portions of an allergen (Rouge et al., 2009), the presence of a large electropositive patch on the Ara h 1 surface could indicate the presence of a relatively high density of epitopes, including either linear or conformational epitopes. We showed the localization of the linear epitopes by mapping the electrostatic potential of the Ara h 1 core monomer. Four epitopes (nos. 9, 14, 19, and 21; Fig. 5B and C) are found in relatively neutral to electropositive regions, while the remaining epitopes are located in neutral to electronegative regions on the Ara h 1 surface. Although detailed data regarding epitopes of the other 7S proteins remain unclear, the data presented here will shed light on the relationship between electrostatic properties and epitopes in 7S globulins.

We also mapped the previously identified epitopes on the Ara h 1 core surface and found that most epitopes either become slightly ($<50\%$ burial; epitopes 8–11, 14–16, 19, and 21) to significantly ($\geq 50\%$) buried (epitopes 7, 13, 17, and 18) upon trimer formation. Only epitope 12 remains uncovered upon trimerization (Table 3; Fig. 4C). The significantly buried epitopes are actually located in regions with high hydrophobicity (Figs. 4A and 5A) and in the N- and C-terminal extension regions, which are points of contact during trimer assembly (Fig. 5C). In the monomeric form, epitope 15 was found to be already significantly buried inside the β -barrel domain (Table 3; Fig. 5C), with 80% of its residues buried. Epitope 15 also has one of the lowest ASA values (194.2 Å²) among the epitopes. In addition, some of the reported residues critical for IgE binding (Burks et al., 1997) are completely inaccessible to

solvent. For example, in the case of epitope 15, none of the five critical residues (H, L, G, F, and G) are exposed on the surface of the molecule. These data indicate that for most of the Ara h 1 epitopes to become completely accessible to immunoglobulins, the allergen has to be completely or partially fragmented, at least in the monomeric form, in the gut.

Our findings may also be very relevant for designing new therapeutic desensitization strategies and even in the diagnosis of peanut allergy, which normally employs only skin and blood testing. Although our knowledge on the localization and accessibility of all epitopes of all allergens, peanut allergens in particular, is not yet complete, this study suggests that the current desensitization approach using intact allergens may be inadequate. The use of both whole allergens and partially degraded components should be seriously considered in the design of future diagnostic and immunotherapeutic strategies.

Acknowledgements

We thank Prof. Reiko Urade (Kyoto University) for the warm encouragement. Diffraction data sets for the crystals were collected from the BL-38B1 station of SPring-8 with approval from the JASRI. This research was supported in part by grants from Asahi Breweries Foundation, the Shimizu Foundation for the Promotion of Immunology, and The Towa Foundation for Food Research (to N.M.). The first author also wishes to thank the Ministry of Education, Culture, Sports, Science, and Technology of Japan for scholarship support.

Accession numbers: The atomic coordinates of the Ara h 1 core region structure have been deposited in the Protein Data Bank under accession code 3SMH.

References

- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D: Biol. Crystallogr.* 66, 213–221.
- Anon, 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D: Biol. Crystallogr.* 50, 760–763.
- Astwood, J.D., Leach, J.N., Fuchs, R.L., 1996. Stability of food allergens to digestion in vitro. *Nat. Biotechnol.* 14, 1269–1273.
- Bindslev-Jensen, C., Briggs, D., Osterballe, M., 2002. Can we determine a threshold level for allergenic foods by statistical analysis of published data in the literature? *Allergy* 57, 741–746.
- Bock, S.A., Munoz-Furlong, A., Sampson, H.A., 2001. Fatalities due to anaphylactic reactions to foods. *J. Allergy Clin. Immunol.* 107, 191–193.
- Boldt, A., Fortunato, D., Conti, A., Petersen, A., Ballmer-Weber, B., Lepp, U., Reese, G., Becker, W.M., 2005. Analysis of the composition of an immunoglobulin E reactive high molecular weight protein complex of peanut extract containing Ara h 1 and Ara h 3/4. *Proteomics* 5, 675–686.
- 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.
- Burks, A.W., Williams, L.W., Helm, R.M., Connaughton, C., Cockrell, G., O'Brien, T., 1991. Identification of a major peanut allergen, Ara h I, in patients with atopic dermatitis and positive peanut challenges. *J. Allergy Clin. Immunol.* 88, 172–179.
- Cabanos, C., Urabe, H., Masuda, T., Tandang-Silvas, M.R., Utsumi, S., Mikami, B., Maruyama, N., 2010. Crystallization and preliminary X-ray analysis of the major peanut allergen Ara h 1 core region. *Acta Crystallogr. F: Struct. Biol. Cryst. Commun.* 66, 1071–1073.
- de Jong, E.C., Van Zijverden, M., Spanhaak, S., Koppelman, S.J., Pellegrini, H., Penning, A.H., 1998. Identification and partial characterization of multiple major allergens in peanut proteins. *Clin. Exp. Allergy* 28, 743–751.
- 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.
- Emmett, S.E., Angus, F.J., Fry, J.S., Lee, P.N., 1999. Perceived prevalence of peanut allergy in Great Britain and its association with other atopic conditions and with peanut allergy in other household members. *Allergy* 54, 380–385.
- Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D: Biol. Crystallogr.* 60, 2126–2132.
- Fukuda, T., Maruyama, N., Salleh, M.R., Mikami, B., Utsumi, S., 2008. Characterization and crystallography of recombinant 7S globulins of Adzuki bean and structure–function relationships with 7S globulins of various crops. *J. Agric. Food Chem.* 56, 4145–4153.
- Grundt, J., Matthews, S., Bateman, B., Dean, T., Arshad, S.H., 2002. Rising prevalence of allergy to peanut in children: data from 2 sequential cohorts. *J. Allergy Clin. Immunol.* 110, 784–789.
- Itoh, T., Garcia, R.N., Adachi, M., Maruyama, Y., Tecson-Mendoza, E.M., Mikami, B., Utsumi, S., 2006. Structure of 8S alpha globulin, the major seed storage protein of mung bean. *Acta Crystallogr. D: Biol. Crystallogr.* 62, 824–832.
- Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637.
- Kleber-Janke, T., Cramer, R., Appenzeller, U., Schlaak, M., Becker, W.M., 1999. Selective cloning of peanut allergens, including profilin and 2S albumins, by phage display technology. *Int. Arch. Allergy Immunol.* 119, 265–274.
- Koppelman, S.J., Knol, E.F., Vlooswijk, R.A., Wensing, M., Knulst, A.C., Hefle, S.L., Gruppen, H., Piersma, S., 2003. Peanut allergen Ara h 3: isolation from peanuts and biochemical characterization. *Allergy* 58, 1144–1151.
- Krisinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D: Biol. Crystallogr.* 60, 2256–2268.
- Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.
- Lee, B., Richards, F.M., 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55, 379–400.
- 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.
- 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.
- Mari, A., Scala, E., Palazzo, P., Ridolfi, S., Zennaro, D., Carabella, G., 2006. Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The allergome platform as a model. *Cell. Immunol.* 244, 97–100.
- Maruyama, N., Adachi, M., Takahashi, K., Yagasaki, K., Kohno, M., Takenaka, Y., Okuda, E., Nakagawa, S., Mikami, B., Utsumi, S., 2001. Crystal structures of recombinant and native soybean beta-conglycinin beta homotrimers. *Eur. J. Biochem.* 268, 3595–3604.
- Maruyama, N., Katsube, T., Wada, Y., Oh, M.H., Barba De La Rosa, A.P., Okuda, E., Nakagawa, S., Utsumi, S., 1998. The roles of the N-linked glycans and extension regions of soybean beta-conglycinin in folding, assembly and structural features. *Eur. J. Biochem.* 258, 854–862.
- Maruyama, Y., Maruyama, N., Mikami, B., Utsumi, S., 2004. Structure of the core region of the soybean beta-conglycinin alpha' subunit. *Acta Crystallogr. D: Biol. Crystallogr.* 60, 289–297.
- Mills, E.N., Jenkins, J., Marigheto, N., Belton, P.S., Gunning, A.P., Morris, V.J., 2002. Allergens of the cupin superfamily. *Biochem. Soc. Trans.* 30, 925–929.
- Osterballe, M., Hansen, T.K., Mortz, C.G., Host, A., Bindslev-Jensen, C., 2005. The prevalence of food hypersensitivity in an unselected population of children and adults. *Pediatr. Allergy Immunol.* 16, 567–573.
- Petrey, D., Honig, B., 2003. GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. *Methods Enzymol.* 374, 492–509.
- Pomes, A., Helm, R.M., Bannon, G.A., Burks, A.W., Tsay, A., Chapman, M.D., 2003. Monitoring peanut allergen in food products by measuring Ara h 1. *J. Allergy Clin. Immunol.* 111, 640–645.
- Roberts, G., 2007. Anaphylaxis to foods. *Pediatr. Allergy Immunol.* 18, 543–548.
- 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.
- Sen, M., Kopper, R., Pons, L., Abraham, E.C., Burks, A.W., Bannon, G.A., 2002. Protein structure plays a critical role in peanut allergen stability and may determine immunodominant IgE-binding epitopes. *J. Immunol.* 169, 882–887.
- Shin, D.S., Compadre, C.M., Maleki, S.J., Kopper, R.A., Sampson, H., Huang, S.K., Burks, A.W., Bannon, G.A., 1998. Biochemical and structural analysis of the IgE binding sites on ara h1, an abundant and highly allergenic peanut protein. *J. Biol. Chem.* 273, 13753–13759.
- 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.
- Sicherer, S.H., Munoz-Furlong, A., Burks, A.W., Sampson, H.A., 1999. Prevalence of peanut and tree nut allergy in the US determined by a random digit dial telephone survey. *J. Allergy Clin. Immunol.* 103, 559–562.
- 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 dial telephone survey: a 5-year follow-up study. *J. Allergy Clin. Immunol.* 112, 1203–1207.
- Skolnick, H.S., Conover-Walker, M.K., Koerner, C.B., Sampson, H.A., Burks, W., Wood, R.A., 2001. The natural history of peanut allergy. *J. Allergy Clin. Immunol.* 107, 367–374.
- Vagin, A., Teplyakov, A., 2010. Molecular replacement with MOLREP. *Acta Crystallogr. D: Biol. Crystallogr.* 66, 22–25.

- Vaguine, A.A., Richelle, J., Wodak, S.J., 1999. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D: Biol. Crystallogr.* 55, 191–205.
- van Boxtel, E.L., Koppelman, S.J., van den Broek, L.A., Gruppen, H., 2008. Determination of pepsin-susceptible and pepsin-resistant epitopes in native and heat-treated peanut allergen Ara h 1. *J. Agric. Food Chem.* 56, 2223–2230.
- van Boxtel, E.L., van Beers, M.M., Koppelman, S.J., van den Broek, L.A., Gruppen, H., 2006. Allergen Ara h 1 occurs in peanuts as a large oligomer rather than as a trimer. *J. Agric. Food Chem.* 54, 7180–7186.
- Wichers, H.J., De Beijer, T., Savelkoul, H.F., Van Amerongen, A., 2004. The major peanut allergen Ara h 1 and its cleaved-off N-terminal peptide; possible implications for peanut allergen detection. *J. Agric. Food Chem.* 52, 4903–4907.