

Isolation and Characterization of Two Complete Ara h 2 Isoforms cDNA

Jean-Marc Chatel^{a,b} Hervé Bernard^b Frank M. Orson^a

^aVeterans Affairs Medical Center, Baylor College of Medicine, Houston, Tex., USA;

^bINRA-Laboratoire d'Immuno-Allergie Alimentaire, Gif-sur-Yvette, France

Key Words

Peanut · Ara h 2 · Isoform · Clones · Allergy

Abstract

Background: Ara h 2 is a major peanut allergen recognized by IgE in more than 90% of patients. After electrophoretic separation the purified protein exists as a doublet, and sequences of one incomplete cDNA and one genomic clone for this allergen have been reported.

Methods: Ara h 2 isoforms were purified and analyzed by mass spectroscopy, and PCR amplification products of Ara h 2 were cloned and sequenced. **Results:** Mass spectroscopy of purified Ara h 2 clearly identified a molecular doublet of 16,670 and 18,050 Daltons. Amplification of a peanut cDNA library using PCR primer pairs located at the amino- and carboxy-terminus revealed 2 bands separated by 50 base pairs, which we cloned and sequenced. Two types of complete cDNA clones were obtained, Ara h 2.01 and Ara h 2.02. Compared to Ara h 2.01 and the previously reported cDNA sequences, Ara h 2.02 is characterized by a 12 amino acid insertion starting at position 75 that contains a third repeat of the major IgE binding epitope DPYSPS. **Conclusion:** We demonstrated the mo-

lecular and genetic characteristics of two Ara h 2 isoforms, revealing that one, Ara h 2.02, might be the more potent allergen.

Copyright © 2003 S. Karger AG, Basel

Introduction

Food allergy poses a considerable public health problem especially in children because of the potential severity of allergic hypersensitivity. In the US, food allergies are the most common cause of anaphylaxis treated in hospital emergency rooms [1]. Peanuts and tree nuts together are the leading cause of fatal and near-fatal reactions [2]. At present, there are no effective therapies for food allergies except avoidance, which can be difficult due to erroneous food labeling, allergen-containing food offered by well-meaning family and friends, and food sharing among children. As a result, accidental ingestion of peanut-containing food is not rare and results in many deaths each year [3].

Seven cDNAs of peanut allergens (Ara h 1 to 7) have been cloned [4–7]. The two major allergens, Ara h 1 and Ara h 2, are recognized by IgE antibodies in the sera from more than 90% of allergic patients. Ara h 2 is a glycopro-

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2003 S. Karger AG, Basel
1018–2438/03/1311–0014\$19.50/0

Accessible online at:
www.karger.com/iaa

Correspondence to: Dr. Jean-Marc Chatel
INRA-Laboratoire d'Immuno-Allergie Alimentaire, CE Saclay, DRM-SPI
Bat 136
F-91191 Gif-sur-Yvette (France)
Tel. +33 1 69 08 45 97, Fax +33 1 69 08 59 07, E-Mail chatel@dsvidf.cea.fr

tein of about 17 kD purified with at least 2 major bands [8, 9]. A partial Ara h 2 cDNA sequence has been determined [5] and a genomic clone isolated [10]. Sequence analysis of Ara h 2 showed homology with seed storage proteins of the conglutin family. At least 10 linear IgE epitopes were identified but three (aa27–36, aa57–66 and aa65–74) were recognized by IgE antibodies in the sera from all patients tested [5]. Epitopes aa57–66 and aa65–74 contained the amino acid sequence DPYSPS that appears to be necessary for IgE binding [5].

In this paper we report the purification of Ara h 2 as two distinct bands, the molecular weight of both as determined by mass spectroscopy, and the complete cDNA sequence for the two corresponding clones obtained by PCR on a cDNA library of *Arachis hypogaea*. Ara h 2.01 is 98% identical to the previously reported incomplete cDNA, and Ara h 2.02 contains an insertion of 12 amino acids starting at position 75, corresponding to the difference in molecular weights of the two bands. In addition, the insertion contains a third repeat of the major linear IgE epitope motif DPYSPS.

Methods

DNA Purification from cDNA Library

A λ gt11 cDNA library constructed with RNA from developing seeds of F78–1339 (*A. hypogaea*), as described [11], was a kind gift of Albert Abbott (Clemson University, S.C., USA). The titer of the cDNA library, $1.8/10^8$ pfu/ml was obtained as described [12]. Fifty thousand pfu were diluted in SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.5, 0.01% gelatin) and added to 500 μ l of *Escherichia coli* MRA previously grown until $OD_{600} = 0.5$ and resuspended in 10 mM $MgSO_4$. After 15 min at 37°C, we added 7.5 ml of LB M/M Top agarose (LB broth, maltose 0.2%, $MgSO_4$ 10 mM, high melting point agarose 0.4%, low melting point agarose 0.3%) and then plated it on 150 mm NZY Agarose (Yeast extract 0.5%, N-Z-Amine A 1%, Casaminoacids 0.1%, NaCl 100 mM, $MgSO_4$ 8 mM, high melting point agarose 0.7%). After an overnight (ON) incubation, at 37°C, 5 ml SM buffer was added on each plate. Plates were shaken at room temperature (RT) for 1 h. We then added 2 ml chloroform for each plate and incubated for 1 h at 37°C with shaking. The slurry was collected and then centrifuged at 15,000 g for 15 min. Supernatant was loaded on gradient 40/5% glycerol in TMG (10 mM Tris-HCl pH 7.5, $MgSO_4$ 5 mM, gelatin 0.01%) and then centrifuged at 28,000 rpm for 2 h. For each 10 plates, the pellet was resuspended in 5 ml TMG buffer. DNase I was added to a final concentration of 1 μ g/ml, and RNase was added to a final concentration of 5 μ g/ml. The suspension was then incubated at 37°C for 30 min. Enzymes were inactivated by the addition of EDTA (final concentration 20 mM) and SDS (final concentration 0.5%) with a 15-min incubation at 70°C. Then proteinase K was added (final concentration 50 μ g/ml) and the solution was incubated at 60°C for 1 h, followed by inactivation at 70°C for 1 h. The solution was extracted with phenol/chloroform until the interface was clear, and the DNA was precipitated with ethanol.

Amplification

Amplification by PCR was done on purified DNA from λ gt11 cDNA. The typical program used was 94°C for 10 min (1 cycle); then 94°C for 1 min, 55°C for 1 min, 72°C for 1 min (25 cycles) and finally 72°C for 10 min (1 cycle). The reaction mix contained 2.5 U of Taq DNA polymerase and 1 $\times$ assay buffer, 25 mM $MgCl_2$, 0.2 mM dNTPs (all from Gibco BRL), 20 pmol of each primer (Sigma-Genosys) and 10 ng of purified DNA in a total volume of 50 μ l. The primers used were gt11F (GGT GGC GAC GAC TCC TGG AGC CCG, the 5' region of the *Eco*RI insertion site of λ gt11), H2F1 (CTC ACC ATA CTA GTA GCC CTC GCC CTT TTC, the starting nucleotide 2 (N-terminus) of Genbank Accession Number L77197), H2R1 (TTA GTA TCT GTC TCT GCC GCC ACT TTC GAC, the reverse sequence starting nucleotide 475 (C-terminus) of Genbank L77197), T7F (TAA TAC GAC TCA CTA TAG GG, the 5' region of pcDNA3 (Invitrogen) insertion site), and BGHr (TAG AAG GCA CAG TCG AGG, the reverse sequence 3' region of pcDNA3 insertion site).

Cloning and Selection of Positive Clones

Cloning was done in pcDNA3.1/V5-His using the pcDNA3.1/V5-His TOPO TA Expression Kit (Invitrogen) as described by supplier. Briefly, 2 μ l of an amplification reaction using primers gt11F and H2R1 and the Expand High Fidelity PCR System (Boehringer) was directly added to 1 μ l of pcDNA3.1/V5-His TOPO vector, 1 μ l of salt solution (1.2 M NaCl, 0.06 M $MgCl_2$) and 2 μ l of H_2O , and then incubated for 30 min at RT. Then 2 μ l of the cloning reaction was transformed in TOP10 chemically competent cells. The transformed cells were spread on ampicillin (50 μ g/ml) plates, and incubated overnight at 37°C.

First, positive clones were selected for the presence of an insert of the right size by PCR on 20 isolated colonies using primer H2R1 and H2F1. Positive clones after the first selection were tested for an insert in the correct orientation for expression by PCR using T7F promoter (5' of the cloning site) primer and H2F1. Positive clones after the second selection were confirmed by another PCR using primers H2F1 and BGH Reverse primer (3' of the cloning site).

Selected clones were amplified on selective medium, and plasmids were purified (Plasmid Mini Kit, Qiagen) and then sequenced (Lone Star Labs Inc., Houston, Tex., USA).

Protein Purification and Mass Spectroscopy

Ara h 2 was purified from defatted whole peanut protein extract by anion exchange chromatography (Mono Q FPLC, Pharmacia) and preparative RP-HPLC (C4, Vydac). The purified Ara h 2 was analyzed by MALDI/TOF MS (Voyager DERP-2) and characterized by determining the 10 amino-terminal amino acids using an ABI automatic sequencer (Applied Biosystems).

Results

Mass Spectroscopy of Purified Ara h 2 Isoforms

After purification from whole peanut protein extract, Ara h 2 was analyzed by mass spectroscopy. Ara h 2 was present as a doublet of 16,670 and 18,050 Daltons (fig. 1). Both bands were further characterized by sequencing the 10 amino-terminal amino acids. The same sequence,

Fig. 1. Mass spectroscopy on Ara h 2 purified from peanut protein extract. m = Mass; z = charge number.

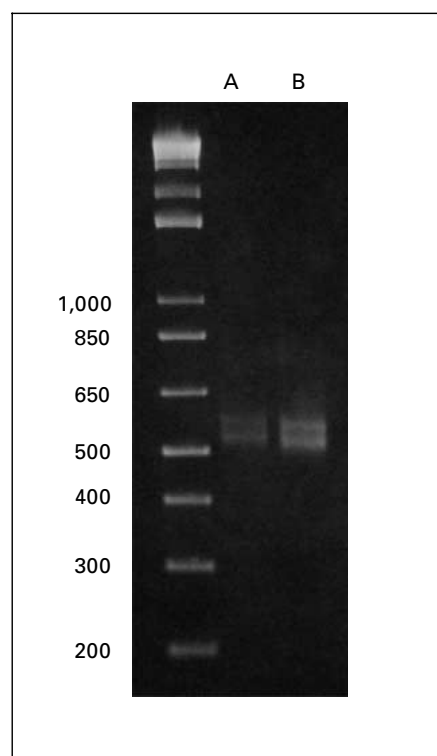
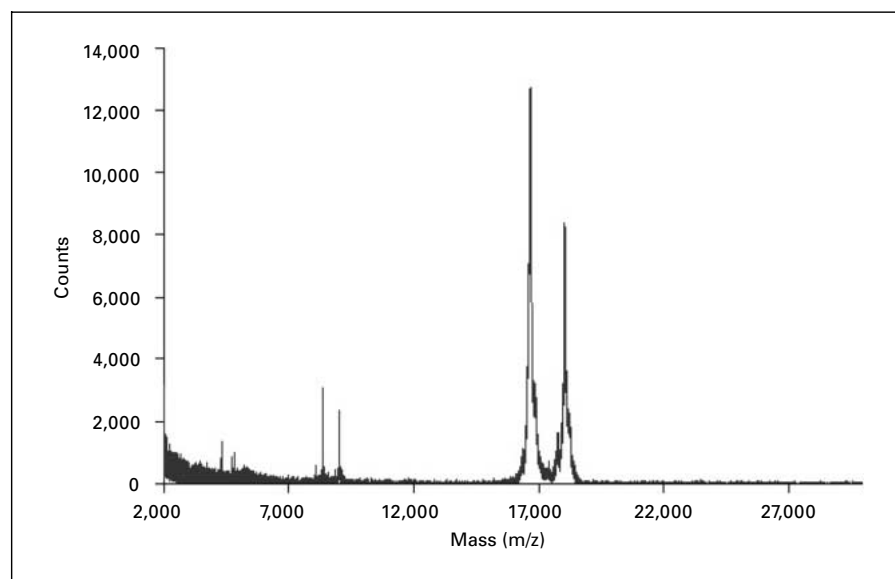


Fig. 2. Lane 1: DNA markers; lane 2: amplification on purified DNA from *A. hypogaeae* λ gt11 cDNA library using gt11F/H2R1 primer pair; lane 3: amplification on purified DNA from *A. hypogaeae* λ gt11 cDNA library using H2F1/H2R1 primer pair.

RQXWELQGDR (X = non-identified), was obtained for each band. This sequence is 100% homologous to the sequences already published [5, 7].

Amplification

PCR using two different primer pairs, gt11F/H2R1 (fig. 2A) and H2F1/H2R1 (fig. 2B), was done on purified DNA from a λ gt11 cDNA library of *A. hypogaeae*. The amplification products were between 500 and 600 bp in length, which is in good agreement with the size of the published cDNA [5, 7]. In both cases, we obtained a doublet with a difference of less than 50 bases in length between the bands indicating the presence of two isoforms of Ara h 2. The difference between these isoforms had to be within the coding region, since second primer pair (fig. 2B) amplified only the coding region of Ara h 2. The amplification bands obtained with the primer pair gt11F/H2R1 were slightly larger than those obtained with H2F1/H2R1 due to the presence of the λ gt11 DNA primer sequences and a short non-coding 5' region.

Cloning

Our first goal was to obtain an Ara h 2 full-length cDNA clone to determine whether these bands represented coding sequence isoforms. We did PCR amplification using the primer pair gt11F/H2R1 and high fidelity DNA polymerase with proofreading activity. A mixture of both bands was cloned into the pcDNA3 expression vector. We selected positive clones by PCR using the

Ara h 2.01	MAKLTILVALALFLLAAHASARQQWELQGDRCQSCLERANLRPCEQHLMQKIQRDEDSY				
Ara h 2.02	MAKLTILVALALFLLAAHASARQQWELQGDRCQSCLERANLRPCEQHLMQKIQRDEDSY				
	10 20 30 40 50 60				
	70 80 90 100				
Ara h 2.01	ER <u>DPYSPS</u> <u>DPYSPS</u> -----PYDRRGAGSSQHQRCCNELNEFENNQRCMCEA				
Ara h 2.02	GR <u>DPYSPS</u> <u>DPYSPS</u> <u>QDPDRRDPYSPS</u> PYDRRGAGSSRHQRCCNELNEFENNQRCMCEA				
	70 80 90 100 110 120				
	110 120 130 140 150 160				
Ara h 2.01	LQQIMENQSDRLQGRQQEQQFKRELRLNLPQQCGLRAPQRCDLVESGGRDRY				
Ara h 2.02	LQQIMENQSDRLQGRQQEQQFKRELRLNLPQQCGLRAPQRCDLVESGGRDRY				
	130 140 150 160 170				

Fig. 3. Sequence comparison between Ara h 2.01 and Ara h 2.02. The amino acid sequences were deduced from nucleotide sequence. The 12 amino acids insertion is in red; the major IgE epitope DPYSPS is **underlined and bold**. ‘.’ = Amino acids are equivalent; ‘:’ = amino acids are the same.

primer pairs H2F1/H2R1, T7F/H2R1 and H2F1/BGHR. Both the sense and antisense strands of clones positive for the two amplification products were sequenced. We obtained clones with inserts of either 545 or 581 bp. Each insert included a 56-bp region coming from λ gt11, a 6 bp 3' non coding region and one of two open reading frames of 483 bp (Ara h 2.01) or 519 bp (Ara h 2.02) starting with an ATG codon and ending with a TAA codon. For both clones, the 5' non-coding region was ACA ACA. In our clone Ara h 2.01 the coding region is slightly longer (plus 9 bp), than the Ara h 2 sequence for Genbank clone L77197. Those 3 aa correspond to the N-terminal sequence MAK. In the overlapping sequence, the two clones are 98.1% identical in amino acids. Ara h 2.02 contains an insertion of 12 aa, QDPDRRDPYSPS, starting at position 75 (fig. 3). This insertion contains a third repeat of the linear IgE recognition motif DPYSPS, which is already, present twice in the published Ara h 2 sequence (Genbank L77197). The calculated difference between the two theoretical protein MW of Ara h 2.01 and Ara h 2.02 is 1,431 Daltons, which is in good agreement with the difference in value obtained by mass spectroscopy on purified protein, 1,380 Daltons.

Discussion

Our work describes for the first time the complete cDNA sequence of Ara h 2.01 and that of a second isoform Ara h 2.02 (Genbank Accession Number

AY158467). This isoform is principally characterized by an insertion of 12 amino acids starting at position 75.

It is well known that Ara h 2 can be purified as a doublet. Burks et al. [8] described Ara h 2 as a doublet with a mean molecular weight of 17 kD and de Jong et al. [9] as a doublet of 14 and 21 kD. Our work confirms the presence of two isoforms of Ara h 2. Using mass spectroscopy, we determined very precisely the MW of both forms. Using 2-D gel analysis, Burks et al. [8] showed that Ara h 2 exists as four distinct isoforms, with the additional two forms presumably reflecting post-translational modifications.

PCR amplification with primers at the ends of Ara h 2 using a cDNA library of *A. hypogaeae*, resulted in two distinct bands. We cloned both bands and obtained the first complete cDNA of Ara h 2 confirming the previous partial cDNA (Genbank clone L77197) sequence and demonstrating the complete cDNA of an isoform of Ara h 2, Ara h 2.02. This isoform is distinguished principally by an insertion of 12 amino acids starting at position 75, containing the motif DPYSPS which is already repeated two times in Ara h 2.01. Two independent clones of the higher MW band from different amplifications of the cDNA library confirmed this sequence difference. This motif has been described as the principal IgE binding epitope of Ara h 2. This hexameric peptide is recognized by serum IgE from all the peanut-hypersensitive patients tested in one study [5], and serum IgE antibodies that recognize this peptide represent the majority of Ara h 2-specific IgE found in these patients. As a result, Ara h 2.02 could potentially be a more powerful allergen than Ara h 2.01.

The comparison of our clones and the genomic clone indicate that the location of ATG codon and 5' coding region not present in the previously reported cDNA sequence (Genbank L77197) match perfectly with the genomic clone [10]. Nevertheless, the 5' non-coding region is completely different, and the 3' coding region after amino acid 156 is also different. No insertion was present in the genomic clone.

In plants, there are two classes of introns: the U2-class, characterized by a GT 5' intron end and AG 3' end and the U12-class, characterized by an AT-AC 5'-3' intron end [13]. Ara h 2.02 insertion does not correspond to either of these classes; therefore, it does not seem coming from a splicing event, implying the existence of a proper genomic clone for Ara h 2.02.

It is important to characterize the different isoforms of an allergen because, as in the case of Bet v 1, the major birch pollen allergen, some of these isoforms may have

differing allergenicity. Bet v 1 shows a considerable degree of heterogeneity among the 13 different cDNA clones coding for isoforms of this allergen [14]. These isoforms show an amino acid identity from 84.4 to 99.4%. Nevertheless, some of these isoforms, like Bet v 1d, have low IgE-binding activity but strong T-cell proliferation activity [15]. Considering those properties, it has been suggested that Bet v 1d could be used in immunotherapy. Bet v 1d also induces different immune responses after genetic immunization than Bet v 1a [16]. Thus, the existence of different isoforms of peanut allergens may have similar potential clinical implications.

Acknowledgement

This study was supported by the Department of Veterans Affairs and the Huffington Foundation, Houston, Tex.

References

- 1 Yocum MW, Butterfield JH, Klein JS, Volcheck GW, Schroeder DR, Silverstein MD: Epidemiology of anaphylaxis in Olmsted County: A population-based study. *J Allergy Clin Immunol* 1999;104:452-456.
- 2 Sicherer SH, Munoz-Furlong A, Burks AW, Sampson HA: Prevalence of peanut and tree nut allergy in the US determined by a random digit dial telephone survey. *J Allergy Clin Immunol* 1999;103:559-562.
- 3 Bock SA, Munoz-Furlong A, Sampson HA: Fatalities due to anaphylactic reactions to foods. *J Allergy Clin Immunol* 2001;107:191-193.
- 4 Burks AW, Shin D, Cockrell G, Stanley JS, Helm RM, Bannon GA: Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. *Eur J Biochem* 1997;245:334-339.
- 5 Stanley JS, King N, Burks AW, Huang SK, Sampson H, Cockrell G, Helm RM, West CM, Bannon GA: Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. *Arch Biochem Biophys* 1997;342:244-253.
- 6 Rabjohn P, Helm EM, Stanley JS, West CM, Sampson HA, Burks AW, Bannon GA: Molecular cloning and epitope analysis of the peanut allergen Ara h 3. *J Clin Invest* 1999;103:535-542.
- 7 Kleber-Janke T, Cramer R, Appenzeller U, Schlaak M, Becker WM: Selective cloning of peanut allergens, including profilin and 2S albumins, by phage display technology. *Int Arch Allergy Immunol* 1999;119:265-274.
- 8 Burks AW, Williams LW, Connaughton C, Cockrell G, TJ OB, Helm RM: Identification and characterization of a second major peanut allergen, Ara h II, with use of the sera of patients with atopic dermatitis and positive peanut challenge. *J Allergy Clin Immunol* 1992;90:962-969.
- 9 de Jong EC, van Zijverden M, Spanhaak S, Koppelman SJ, Pellegrini H, Penninks AH: Identification and partial characterization of multiple major allergens in peanut proteins. *Clin Exp Allergy* 1998;28:743-751.
- 10 Viquez OM, Summer CG, Dodo HW: Isolation and molecular characterization of the first genomic clone of a major peanut allergen, Ara h 2. *J Allergy Clin Immunol* 2001;107:713-717.
- 11 Jung S, Swift D, Sengoku E, Patel M, Teule F, Powell G, Moore K, Abbott A: The high oleate trait in the cultivated peanut (*Arachis hypogaea* L.). I. Isolation and characterization of two genes encoding microsomal oleoyl-PC desaturases. *Mol Gen Genet* 2000;263:796-805.
- 12 Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, ed 2. Cold Spring Harbor, Cold Spring Harbor Laboratory Press, 1989.
- 13 Lorkovic ZI, Wiczeorek Kirk DA, Lambrmon MH, Filipowicz W: Pre-mRNA splicing in higher plants. *Trends Plant Sci* 2000;5:160-167.
- 14 Swoboda I, Jilek A, Ferreira F, Engel E, Hoffmann-Sommergruber K, Scheiner O, Kraft D, Breiteneder H, Pittenauer E, Schmid E: Isoforms of Bet v 1, the major birch pollen allergen, analyzed by liquid chromatography, mass spectrometry, and cDNA cloning. *J Biol Chem* 1995;270:2607-2613.
- 15 Ferreira F, Hirtenlehner K, Jilek A, Godnik-Cvar J, Breiteneder H, Grimm R, Hoffmann-Sommergruber K, Scheiner O, Kraft D, Breitenbach M, Rheinberger HJ, Ebner C: Dissection of immunoglobulin E and T lymphocyte reactivity of isoforms of the major birch pollen allergen Bet v 1: Potential use of hypoallergenic isoforms for immunotherapy. *J Exp Med* 1996; 183:599-609.
- 16 Hartl A, Kiesslich J, Weiss R, Bernhaupt A, Mostböck S, Scheiblhofer S, Flockner H, Sippl M, Ebner C, Ferreira F, Thalhamer J: Isoforms of the major allergen of birch pollen induce different immune responses after genetic immunization. *Int Arch Allergy Immunol* 1999;120: 17-29.