

Recombinant Peanut Allergen *Ara h I* Expression and IgE Binding in Patients with Peanut Hypersensitivity

A. Wesley Burks,* Gael Cockrell,* J. Steven Stanley,† Ricki M. Helm,* and Gary A. Bannon†

Departments of *Pediatrics and †Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences and Arkansas Children's Hospital Little Rock, Arkansas 72205

Abstract

Peanut allergy is a significant health problem because of the frequency, the potential severity, and the chronicity of the allergic sensitivity. Serum IgE from patients with documented peanut hypersensitivity reactions and a peanut cDNA expression library were used to identify clones that encode peanut allergens. One of the major peanut allergens, *Ara h I*, was selected from these clones using *Ara h I* specific oligonucleotides and polymerase chain reaction technology. The *Ara h I* clone identified a 2.3-kb mRNA species on a Northern blot containing peanut poly (A)⁺ RNA. DNA sequence analysis of the cloned inserts revealed that the *Ara h I* allergen has significant homology with the vicilin seed storage protein family found in most higher plants. The isolation of the *Ara h I* clones allowed the synthesis of this protein in *E. coli* cells and subsequent recognition of this recombinant protein in immunoblot analysis using serum IgE from patients with peanut hypersensitivity. With the production of the recombinant peanut protein it will now be possible to address the pathophysiologic and immunologic mechanisms regarding peanut hypersensitivity reactions specifically and food hypersensitivity in general. (*J. Clin. Invest.* 1995; 96:1715–1721.) Key words: food hypersensitivity • peanuts • legumes • vicilins • epitopes

Introduction

Peanut allergy is a significant health problem because of the potential severity of the allergic reaction, the chronicity of the allergic sensitivity, and the ubiquity of peanut products. Peanuts, fish, tree nuts, and shellfish account for the majority of food hypersensitivity reactions in adults, while peanuts, milk, and eggs cause > 80% of food hypersensitivity reactions in children (1, 2). Peanut hypersensitivity reactions often tend to be quite severe in nature, sometimes resulting in episodes of fatal anaphylaxis (3, 4). Unlike the food hypersensitivity reactions to milk and eggs, peanut hypersensitivity reactions usually persist into adulthood and last for a lifetime (5). Despite the significant prevalence of peanut hypersensitivity reactions and several fatalities annually, the identification of the clinically relevant anti-

gens and an understanding of the immunobiology of peanut hypersensitivity is just beginning (6).

Recently a number of allergens have been identified which stimulate IgE production and cause IgE mediated disease (7). The identification and purification of these allergens is essential for the immunological studies necessary to understand their role in stimulating IgE antibody formation. Significant information now exists on the identification and purification of inhaled allergens from pollens, dust mites, animal danders, insects, and fungi (7). By comparison, few food allergens known to cause IgE-mediated reactions have been similarly identified and purified (8–10).

Recombinant methodology to clone allergens provides an efficient means of producing pure polypeptides which, in their native source, form complex mixtures and are often represented in only very small amounts. Several inhaled allergens have been cloned, including the allergens of house dust mites (11) and pollen grains (12), in comparison little work has been directed toward producing recombinant food allergens.

Because of the prevalence and severity of peanut hypersensitivity reactions in both children and adults, coupled with the recent identification of two major peanut allergens that are involved in this process (6, 13), we set out to clone and characterize the *Ara h I* peanut allergen. In this communication we report the isolation and characterization of the *Ara h I* mRNA and the identification of this allergen as belonging to the vicilin seed storage protein family. In addition, this recombinant allergen can be produced in *E. coli* cells and is recognized by serum IgE from patients with documented peanut hypersensitivity reactions.

Methods

Patients. Serum from eighteen patients with documented peanut hypersensitivity (mean age, 25 yr) was used to identify peanut allergens. Each of these individuals had a positive immediate prick skin test to peanut and either a positive double blind, placebo controlled, food challenge (DBPCFC) or a convincing history of peanut anaphylaxis (laryngeal edema, severe wheezing, and/or hypotension). One individual with elevated serum IgE levels (who did not have peanut specific IgE or peanut hypersensitivity) was used as a control in these studies. Details of the challenge procedure and interpretation have been discussed previously (6). At least five mls of venous blood were drawn from each patient and allowed to clot, and the serum was collected. All studies were approved by the Human Use Advisory Committee at the University of Arkansas for Medical Sciences.

Isolation and amino acid sequence analysis of peanut allergen *Ara h I*. *Ara h I* was purified to near homogeneity from whole peanut extracts according to the methods of Burks et al (6). Purified *Ara h I* was electrophoresed on 12.5% acrylamide mini-gels (Bio-Rad, Hercules, CA) in Tris glycine buffer. The gels were stained with 0.1% Coomassie blue in 10% acetic acid, 50% methanol, and 40% water for 3 h with continuous shaking. Gel slices containing *Ara h I* were sent to the W. M. Keck Foundation (Biotechnology Resource Laboratory, Yale University, New Haven, CT) for amino acid sequencing. Initial sequencing

Address correspondence to A. Wesley Burks, M.D., Arkansas Children's Hospital, 800 Marshall St., Little Rock, AR 72202. Phone: 501 320-1060; FAX: 501 320-3551.

Received for publication 18 July 1994 and accepted in revised form 8 June 1995.

J. Clin. Invest.

© The American Society for Clinical Investigation, Inc.

0021-9738/95/10/1715/07 \$2.00

Volume 96, October 1995, 1715–1721

indicated that the amino terminal end of *Ara h I* was blocked. In order to obtain protein sequencing data *Ara h I* was treated with trypsin and peptides were selected for further analysis. Amino acid sequencing of tryptic peptides was performed on an Applied Biosystems sequencer with an on-line HPLC column that was eluted with increasing concentrations of acetonitrile.

Peanut RNA isolation and northern (RNA) gels. Three commercial lots from the 1979 crop of medium grade peanut species, *Arachis hypogaea* (Florunner) were obtained from North Carolina State University for this study. Total RNA was isolated from one gram of this material according to procedures described by Larsen (14). Poly(A)⁺ RNA was isolated using a purification kit supplied by Collaborative Research (Bedford, MA) according to manufacturer's instructions. Poly(A)⁺ RNA was subjected to electrophoresis in 1.2% formaldehyde agarose gels, transferred to nitrocellulose, and hybridized with ³²P-labeled probes according to the methods of Bannon et al (15).

cDNA expression library construction and screening. Peanut poly(A)⁺ RNA was used to synthesize double-stranded cDNA according to the methods of Watson and Jackson (16) and Huynh et al. (17). The cDNA was treated with EcoRI methylase and then ligated with EcoRI and XhoI linkers. The DNA was then ligated with EcoRI-XhoI cut, phosphatase treated Lambda ZAP XR phage arms (Stratagene, LaJolla, CA) and in vitro packaged. The library was 95% recombinants carrying an average insert size of > 400 bp as determined by sizing of randomly selected clones. The library was screened using an IgE antibody pool consisting of an equal volume of serum from each patient with peanut hypersensitivity. Detection of the primary antibody was either with alkaline phosphatase labeled anti-IgE or I¹²⁵-labeled anti-IgE antibody performed according to manufacturer's instructions. Positive plaques were subjected to subsequent screens using the same pooled serum until all nonreacting plaques were removed. The remaining positive plaques were then rescreened with serum from a patient with elevated total serum IgE who did not have peanut specific IgE to ensure that we were not isolating non-specific, IgE binding clones.

PCR amplification of the *Ara h I* mRNA sequence. Using the oligonucleotide GA(TC)AA(AG)GA(TC)AA(TC)GTNAT(TCA)GA(TC)CA derived from amino acid sequence analysis of the *Ara h I* (63.5 kD) peanut allergen as one primer and a 27 nucleotide long oligo dT stretch as the second primer a portion of the nucleotide sequence that encodes this protein was amplified from peanut cDNA. Reactions were carried out in a buffer containing 3 mM MgCl₂, 500 mM KCl, 100 mM Tris-HCl, pH 9.0. Each cycle of the polymerase chain reaction consisted of 1 min at 94°C, followed by 2 min at 42°C, and three minutes at 72°C. Thirty cycles were performed with both primers present in all cycles. From this reaction a 400-bp fragment was amplified and subsequently cloned into a TA vector by standard protocols (Promega, Madison, WI).

DNA sequencing and analysis. Sequencing was done according to the methods of Sanger et al. (18) using a series of clones constructed by ExoIII digestion of the original DNA isolate or oligonucleotide primers directed to different regions of the clone. Sequence analysis was done on the University of Arkansas for Medical Science's Vax computer using the Wisconsin DNA analysis software package.

Production of recombinant *Ara h I* protein. The *Ara h I* cDNA was ligated into the EcoRI site of a pBluescript vector (Stratagene, LaJolla, CA). This vector contains 111 nucleotides of the Beta galactosidase gene before the EcoRI site. When *E. coli* JM109 cells carrying this construct are induced with IPTG they produce a fusion protein consisting of 37 amino acids derived from Beta galactosidase followed by the *Ara h I* protein. Exponentially growing cells are induced with 1 mM IPTG for 4 h at 37°C. Cells are then pelleted and resuspended in SDS-sample buffer, placed in a boiling water bath for 5 min and then either used immediately for immunoblot analysis or stored at -20°C until needed.

IgE immunoblot analysis. SDS-PAGE was performed by the method of Laemmli (19). All gels were composed of a 10% acrylamide resolving gel and 4% acrylamide stacking gel. Electrophoretic transfer and immunoblotting on nitrocellulose paper were performed by the procedures of Towbin et al. (20). The blots were incubated with antibodies diluted in a solution containing TBS and 1% bovine serum albumin for

Table I. Amino Acid Sequence of *Ara h I* Peptides

Peptide	Amino acid sequence
I	I-F-L-A-G-D-K-D-N-V-I-D-Q-I-E-K
II	K-G-S-E-E-E-G-D-I-T-N-P-I-N-L-R
III	N-N-P-F-Y-F-P-S-R-R

The amino acid sequence of three tryptic peptides derived from purified *Ara h I* protein was determined. The sequence is shown as the one letter amino acid code.

at least 12 h at 4°C or for 2 h at room temperature. Detection of the primary antibody was with ¹²⁵I-labeled anti-IgE antibody.

Results

Isolation and partial amino acid sequence determination of peptides derived from the *Ara h I* protein. Purified *Ara h I* protein was treated with trypsin and the peptide products separated from one another by high performance liquid chromatography. Three peptide fractions, selected on the basis of their separation from each other and other fractions in the mix, were used for amino acid sequence determination. During the course of sequencing it was noted that fraction I and III consisted of a single peptide species (peptide I and peptide III, respectively). Fraction II consisted of one major peptide (peptide II) with numerous minor peptide contaminants which complicated sequence determination. However, it was possible to determine the first 16 residues of the major peptide in fraction I and II and the first 10 residues of the major peptide in fraction III. The amino acid sequence determined for each peptide is noted in Table I.

Isolation of clones that produce antigens recognized by peanut-specific IgE from patients with peanut hypersensitivity. RNA isolated from the peanut species, *Arachis hypogaea* (Florunner) was used to construct an expression library for screening with serum IgE from patients with peanut hypersensitivity. Numerous IgE-binding clones were isolated from this library after screening 10⁶ clones with serum IgE from a pool of patients with reactivity to most peanut allergens by western blot analysis. Since the number of plaques reacting with serum IgE was too large to study all in detail we randomly selected a small portion of the positive plaques for further purification. Phage positive for IgE binding were plaque purified to homogeneity and then tested for their ability to react with serum IgE collected from a patient without peanut hypersensitivity. All of the selected clones were intensely positive when incubated with serum IgE from patients with peanut hypersensitivity. In contrast, these same clones did not react with control serum IgE. These results show that we have isolated numerous clones capable of producing IgE recognizable antigens specific to patients who have peanut hypersensitivity.

Identification and characterization of clones that encode peanut allergen *Ara h I*. To help identify which of the many IgE positive clones encoded the *Ara h I* allergen, a hybridization probe was constructed using an oligonucleotide developed from *Ara h I* amino acid sequence and PCR technology. The oligonucleotide sequence GA(TC)AA(AG)GA(TC)AA(TC)GTN-AT(TCA)GA(TC)CA was derived from amino acid residues located within peptide I (Table I) of the *Ara h I* peanut allergen. Utilizing this oligonucleotide as one primer and a 27-nucleotide

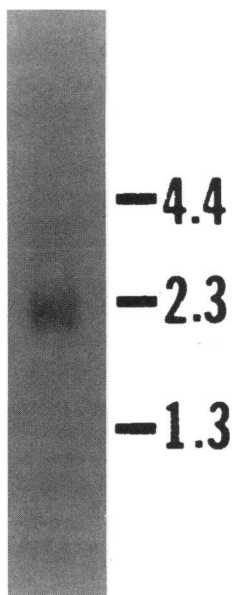


Figure 1. An *Ara h I* clone hybridizes to a 2.3-kb peanut mRNA. Peanut poly (A)⁺ RNA was isolated from *Arachis hypogaea* (Florunner) species and 10 μ g were electrophoresed on denaturing formaldehyde agarose gels. Insert from 41B was purified, labeled with alpha-³²P-dCTP, and used as a hybridization probe of a Northern blot of this gel. Sizes of known RNA species are expressed in kilobases along the right side of the figure.

oligo dT stretch as the second primer a portion of the mRNA sequence that encodes this protein was amplified from peanut cDNA. This 400-bp DNA fragment was subsequently cloned and sequenced by the Sanger dideoxy (18) method. DNA sequence analysis revealed that the 400-bp DNA fragment contained a poly A stretch on one end and the *Ara h I* specific nucleotide sequence on the other end. In addition, this clone contained nucleotide sequence correctly encoding the remaining carboxy terminal portion of peptide I. Thus, an *Ara h I* specific clone has been isolated and it can be used as a hybridization probe to identify which of the many IgE positive clones selected encodes the *Ara h I* allergen.

We hybridized a Southern blot containing four of the IgE selected cloned DNAs with a ³²P-labeled, *Ara h I* PCR amplification product to determine which of the isolated clones encoded the *Ara h I* peanut allergen. All of the clones were positive for hybridization with this probe. In addition we screened 200,000 clones from the peanut cDNA library using ³²P-labeled *Ara h I* clone as a probe. From this screen, over 100 *Ara h I* positive clones were identified (data not shown). These results indicate that the mRNA encoding the *Ara h I* allergen is an abundant message within this library.

To determine what size mRNA these clones identify, ³²P-labeled insert from one of the largest cDNA clones (41B) was used as a hybridization probe of a Northern blot containing peanut poly(A)⁺ RNA (Fig. 1). This insert hybridized to an ~ 2.3-kb mRNA indicating that this insert probably represented the entire mRNA.

Peanut allergen Ara h I is a vicilin-like seed storage protein. The primary DNA sequence of two of the largest cDNA clones selected (41B and P17) was determined by Sanger dideoxy sequencing using oligonucleotide primers directed to different regions on the insert or a series of subclones constructed by ExoIII digestion of the inserts. Clone 41B carried a 2,050-base insert while clone P17 carried a 1,972 base insert. The first ATG protein synthesis start codon was located at nucleotide position 50–53. The sequence around this codon agrees with the translation initiation sequence found in most eukaryotic mRNAs (21). Each of the inserts contained a large open reading frame starting with this codon and ending with a TGA codon at nucleo-

tide position 1928–1930. Overall, there was > 97% DNA sequence homology between the two inserts.

Both clones were capable of encoding a protein of ~ 68 kD. The amino acid sequence that was determined from *Ara h I* peptides I and III is found in both of these clones. The only difference between the derived and predicted amino acid sequence of both clones occurs at position seven of peptide II in clone P17. At this position there is a glycine residue in the peptide that is missing in the P17 DNA sequence. In addition, both proteins have a signal peptide at the amino terminus (22) and a single glycosylation addition site (NAS) at amino acid position 521–523. These data confirm and extend our conclusion that these clones encode the *Ara h I* allergen.

A search of the GenBank database revealed significant sequence homology between the *Ara h I* cDNA clones and a class of seed storage proteins called vicilins. There was 60–65% homology over > 750 bases when the *Ara h I* DNA sequences were compared with the broad bean and pea vicilins (Table II). These results indicate that the *Ara h I* allergen belongs to a vicilin-like multi-gene family encoding very similar but not identical proteins.

Recognition of recombinant Ara h I by patient sera in an IgE immunoblot assay. IgE immunoblot analysis was initially performed using serum IgE from a pool of patients with peanut hypersensitivity to determine the molecular weight of the recombinant protein and the specificity of the IgE recognition reaction. Fig. 3 (lanes A and B) shows that the IgE pool recognized whole peanut extract and purified native *Ara h I* protein as expected, but did not react with any proteins from an *E. coli* lysate that was prepared from cells carrying vector alone (Fig. 3, lane E). However, instead of the IgE pool recognizing a 68-kD protein produced from clone P17, an unexpectedly small protein was identified (Fig. 3, lane C). On further analysis, we noted that by eliminating the first 93 bases (31 amino acids, 5% of *Ara h I*) of this clone we could produce full length *Ara h I* protein (68 kD) with numerous truncated products that migrated as smaller IgE reactive peptides (Fig. 3, lane D). The presence of truncated *Ara h I* products could be the result of inefficient translation of the amino terminal portion of this protein (23, 24) caused by rare codons, numerous cysteine residues, or secondary structure of the mRNA.

Fig. 4 shows eighteen immunoblot strips of recombinant *Ara h I* (upper panel) or native *Ara h I* (lower panel) that have been incubated with different patient sera. 94% (17/18) of the patients that showed IgE binding to the native allergen also showed some level of binding to the recombinant *Ara h I* protein. Of the 18 patient sera tested in this manner there were varying intensities of IgE binding to the recombinant and native allergen. In general, there was good agreement between the level of IgE binding of recombinant and native *Ara h I* for any individual patient. For example, patients who had high levels of IgE which bound native protein (Fig. 4, lower panel, lanes A–F) also showed high immunoreactivity with recombinant *Ara h I* protein (Fig. 4, upper panel, lanes A–F). Patients who had low levels of IgE which bound native allergen (Fig. 4, lower panel, lanes L–R) showed low reactivity with the recombinant protein (Fig. 4, upper panel, lanes L–R). One peanut sensitive individual (lane K) who had serum specific IgE to native *Ara h I* had no detectable IgE which recognized the recombinant protein (Fig. 4, upper panel, lane K). The differences we have noted between peanut hypersensitive patients could be due to the amount of peanut-specific IgE in individual patients, differ-

1 ATAATCATATATATTATCAATCATCTATATAAGTAGTAGCAGGACGA
.....
50 ATGAGAGGGAGGGTTTCTCCACTGATGCTGTGCTAGGATCCCTTGTCTCG
.....
M R G R V S P L M L L L G I L V L 17
101 GCTTCAGTTTCTGCAACGCATGCCAAGTCACTACCTTACAGAAGAAAC
.....G.....C-----TT.CCG.....
A S V S A T H A K S S P Y Q K K T 34
151 AGAGAACCCTCGCCAGAGGTGCCTCCAGAGTTGTCAACAGGAACCGG
.....
E N P C A Q R C L Q S C Q Q E P 50
201 ATGACTTGAAGCAAAAGGCATGCGAGTCTCGTGCACCAAGTCTCGAGTAT
.....
D D L K Q K A C E S R C T K L E Y 67
251 GATCCTCGTTGTCTATGATCTCCGAGGACACTGGCACCACCAACCA
.....G.....
D P R C V Y D P R G H T G T T N Q 84
301 ACGTTCCCTCCAGGGGACGGACACGTGGCCCAACCCGGAGACTACG
.....CA.....
R S P P G E R T R G R Q P G D Y 100
351 ATGATGACCGCCGTCACCCCAAGAGAGGAAGAGGCCGATGGGACCA
.....
D D D R R Q P R R E E G G R W G P 117
401 GCTGGACCGAGGAGCGTGAAGAGAAAGACTGGAGACAACCAAGAGA
.....A.....
A G P R E R E R E E D W R Q P R E 134
451 AGATTGGAGGGACCAAGTCACTCAGCAGCCAGGAAATAAGCCCGAAG
.....
D W R R P S H Q Q P R K I R P E 150
501 GAAGAGAAGGAGAACAAGAGTGGGGAACACAGGTAGCCATGTGAGGAA
.....G.G.....
G R E G E Q E W G T P G S H V R E 167
551 GAAACATCTCGGAACAACCTTCTACTTCCCGTCAAGGCGTTTATGAC
.....A.....
E T S R N N P F Y F P S R R F S T 184
601 CCGCTACGGGAACCAAAACCGTAGGATCCGGTCTCGACAGGTTTAC
.....C.....
R Y G N Q N G R I R V L Q R F D 200
651 AAAGTCAAGGCAGTTTCCAGAACTCCAGAACTACCGTATTGTGCGATC
.....A.....
Q R S R Q F Q N L Q N H R I V Q I 217
701 GAGGCAAACTAACAATCTTGTCTTCCCAAGCAGCTGATGCTGATAA
.....G.....
E A K P N T L V L P K H A D A D N 234
751 CATCCTTGTATCCAGCAAGGGCAAGCCACCGTAGCAATGGCA
.....A.....
I L V I Q Q G Q A T V T V A N G 250
801 ATAACAGAAAGAGCTTAACTTGTGACAGGGCCATGCACTCAGAAATCCA
.....
N N R K S F N L D E G H A L R I P 267
851 TCCGTTTCATTTCTACATCTGAACCCATGACCAACAGAACCTCAG
.....T.A.....
S G F I S Y I L N R H D N Q N L R 284
901 AGTAGTAAATCTCCATGCCGTTAACAACCCGCGCAGTTTGGAGATT
.....G.....
A A K I S M P V N T P G Q F E D 300

951 TCTTCCGGGAGCAGCCGAGACCAATCATCTACTTGCAGGGCTTCAGC
.....A.....
F F P A S S R D Q S S Y L Q G F S 317
1001 AGGAATACGTTGGAGGCCCTTCAATGCGGAATTCATGAGATACGGAG
.....T.....
R N T L E A A F N A E F N E I R R 334
1051 GGTGTGTTAGAAGAAATGCAGGAGTGAGCAAGAGGAGAGGGGAGA
.....A.....
V L L E E N A G G E Q E E R G Q 350
1101 GGCGATGGAGTACTCGGAGTAGTGAGAAATGAAGGAGTATAGTCAA
.....C-----T.....
R R W S T R S S E N N E G V I V K 367
1151 GTGTCAAAGGAGCAGCTTGAAGAACTACTAAGCAGCGTAAATCCGCTC
.....C.....
V S K E H V E E L T K H A K S V S 384
1201 AAAGAAAGGCTCCGAAGAAGAGGAGATATCACCACCAATCAACTGA
.....A.....
K K G S E E E G D I T N P I N L 400
1251 GAGAAGGCGAGCCGATCTTTCTAACTTTGGGAAATTTATTTAGGTTG
.....T.....G.....
R E G E P D L S N N F G K L F E V 417
1301 AAGCCAGACAAGAAGAACCCAGCTTCAGGACTGGACATGATGCTCAC
.....
K P D K K N P Q L Q D L D M M L T 434
1351 CTGTGTAGAGATCAAGAAGGAGCTTTGATGCTCCACACTCAACTCAA
.....
C V E I K E G A L M L P H F N S 450
1401 AGGCCATGGTTATCGTCGTCGCAACAAGGAAGTGGAAACCTTGAAC
.....C.....
K A M V I V V V N K G T G N L E L 467
1451 GTGGCTGAAGAAAAGAGCAACAACAGAGGGGACGGCGGAA-----
.....A.....CAAGAGTC
V A V R K E Q Q Q R G R R E - - - 481
1493 -GAAGAGGAGGACGAAGACGAAGAAGAGGAGGAAGTAAACAGAGGTGC
G.....A.....G.....T.....
E E E D E D E E E E G S N R E V 497
1542 GTAGTACACAGCGAGTTGAAGGAAGCGATGTTTCATCATGCCAGCA
.....T.....
R R Y T A R L K E G D V F I M P A 514
1592 GCTCATCCAGTAGCCATCAACGCTTCTCCGAACTCCATCTGCTGGCTT
.....
A H P V A I N A S S E L H L L G F 531
1642 CGGTATCAACGCTGAAACAACCAAGAACTTCTTGCAGGTGATAAGG
.....
G I N A E N N H R I F L A G D K 547
1692 ACAATGTGATAGACCAGATAGAGAAGCAAGCAAGGATTTAGCATCCCT
.....
D N V I D Q I E K Q A K D L A F P 564
1742 GGTCCGGTGAACAAGTTGAGAAGCTCATCAAAACAGAAAGGAAATCCA
..T.....G.....G.....
G S G E Q V E K L I K N Q K E S H 581
1792 CITTTGAGTGTCTGCTCAATCTCAATCTCAATCTCCGTCGCTCCTG
.....CG.....
F V S A R P Q S Q S Q S P S S P 597
1842 AGAAAGAGTCTCTGAGAAAGAGATCAAGAGGAGGAAAACCAAGGAGG
-----A.....
E K E S P E K E D Q E E E N Q G G 614
1892 AAGGGTCCACTCCTTTCAATTTTGAAGGCTTTTAACTGAGAATGGAGGCA
.....A.....
K G P L L S I L K A F N 626
1942 ACTTGTATGTATCGATAATAAGATCAGCCTTTTGTACTCTACTATCCAA
.....C.....A.....
1992 AACTTATCAATAAATAAAACGTTTGTGCGTTGTTTCTCCAAAAA
.....

Figure 2. Nucleotide sequence of an *Ara h I* cDNA clone. The nucleotide sequence of clone 41B is shown on the first line. The second line depicts clone P17 DNA sequence with dots (.) representing nucleotides that are the same, dashes (-) nucleotides that are missing, and A, C, G, or T nucleotides that differ between the two DNA sequences. The protein synthesis start (ATG) and stop (TGA) sites are underlined along with a consensus polyadenylation signal (AATAAA). Bold amino acid residues are those areas which correspond to the determined amino acid sequence of peptides I, II, and III of *Ara h I* (Table I). The numbers on the left of the figure indicate the nucleotide sequence, and those on the right correspond to the deduced amino acid sequence. These sequence data are available from GenBank under accession number L34402.

Table II. Homology of the *Ara h I* Gene to Plant Vicilins

	Clone 41B		Clone P17	
	bp overlap	Percent homology	bp overlap	Percent homology
Broad bean	1,081	64.3	985	62.3
Pea	1,078	64.2	961	62.5
Soybean	323	65.9	815	61.2

The Wisconsin DNA analysis software package was used to search for homology between the *Ara h I* nucleotide sequence and any DNA sequence contained in the data base. Significant homology was observed between *Ara h I* and the plant vicilins.

ences in affinity of patient-specific IgE for peanut, or that some patients recognize only certain peanut proteins.

Discussion

Peanuts are one of the most allergenic foods (25). Sensitive individuals may experience symptoms ranging from urticaria to anaphylaxis (25). Multiple cases of fatal anaphylaxis have been reported (4). Peanut sensitivity appears early in life and often persists indefinitely. Because of the significance of the allergic reaction and the widening use of peanuts as protein extenders in processed foods, the risk to the peanut-sensitive individual is increasing.

Various studies over the last several years have examined the nature and location of the multiple allergens in peanuts (26). Taylor et al. demonstrated that the allergenic portion of peanuts was in the protein portion of the cotyledon (27). Our laboratory recently identified two major allergens from peanut extracts, designated *Ara h I* (6) and *Ara h II* (13). Greater than 90% of our patients who were challenge positive to peanut had specific IgE to these proteins.

The *Ara h I* nucleotide sequences identified in this report have significant sequence homology with the vicilin family of seed storage proteins of other legumes (soybean, pea, common bean, etc.). The major seed storage proteins of legumes are globulins that are represented in most legumes by two different types of polypeptides, the nonglycosylated legumins and glycosylated vicilins. The genes for the glycosylated seed storage proteins of higher plants code for proteins that are classified by their size into small (50 kD) and large (70 kD) vicilins (28). A comparison of the vicilin amino acid sequences reveals considerable amino acid homology between the small and large vicilins in the carboxy terminal portion of these molecules. The major difference between the large and small vicilin preproteins is the existence of an additional tract of amino acids at the amino terminal end of the large vicilins (29). The information generated in our laboratory demonstrating that the major peanut allergens are vicilin-like proteins may explain why patients with peanut hypersensitivity and peanut-specific IgE tend to have serum IgE to multiple other legume proteins. Since the vicilins of most major plants share significant sequence homology in their carboxy terminal portion, it is not surprising that serum specific IgE would tend to bind to several vicilin proteins from different sources. However, despite patients with legume hypersensitivity having IgE to multiple legume proteins (peanuts, soybeans, peas, etc.) they generally have clinical food hypersensitivity to only one food in the legume family. Because the

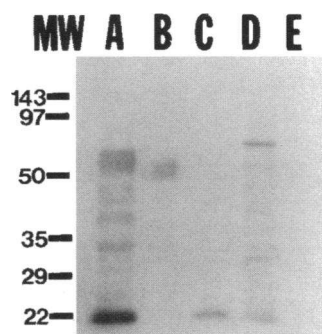


Figure 3. Serum IgE from a pool of peanut hypersensitive patients recognize recombinant *Ara h I*. Serum IgE from a pool of patients with reactivity to most peanut allergens was used to detect whole peanut extract (lane A), purified native *Ara h I* (lane B), recombinant *Ara h I* (lanes C and D), or *E. coli* extract (lane E). In preliminary experiments, we noted that full length clone P17 made small quantities

of truncated recombinant protein (lane C). We found that by eliminating the first 93 bases of this clone we could produce large quantities of nearly full length *Ara h I* protein (lane D). The recombinant *Ara h I* (lane D; 68 kD) is larger than the native *Ara h I* (lane B; 65 kD) because the recombinant protein includes 37 amino acids of Beta galactosidase (See Methods for details of the expression system). Note that the serum IgE pool does not recognize any proteins in the *E. coli* extract (lane E) and therefore the other bands in lane D are truncated versions of *Ara h I*. MW, molecular weight standards expressed in kilodaltons.

amino terminal domains of the large glycosylated (vicilin) proteins share little or no homology, the immune response to this portion of the protein may be responsible for the severe and chronic hypersensitivity response characteristic of peanuts.

We have demonstrated that the cloned *Ara h I* gene is capable of producing a protein product in procaryotic cells that is recognized by serum IgE from a large proportion of individuals with documented peanut hypersensitivity. These results are significant in that they indicate that some of the allergenic epitopes responsible for this reaction are linear amino acid sequences that do not include a carbohydrate component. These findings may provide the basis for the improving diagnosis and therapy of persons with food hypersensitivity.

Current diagnosis of food hypersensitivity relies on a significant clinical history plus evidence of specific IgE to the food allergen in question. The absence of specific IgE to a food means there is a > 95% probability that the ingestion of the food will not lead to clinical symptoms. However, the presence of specific IgE to a particular food has only at best a 50% positive predictive value when correlated with a positive food challenge (1). One explanation of this low predictive value is the current use of crude allergen mixtures. The use of recombinant allergens should make it possible to obtain diagnostics with optimal concentrations of each allergen or to develop panels of mixtures of large numbers of recombinant allergens and then test for individual components. The disadvantages of recombinant allergens are that the IgE binding capacity of some may not be that of the corresponding natural allergen and that the number of recombinant allergens to be produced for any one food may need to be quite large.

Another possible use for recombinant peanut allergens is in immunotherapy. Allergen immunotherapy is an effective therapeutic modality for patients with insect sting hypersensitivity when they have experienced significant systemic symptoms (30). Because allergen immunotherapy can down-regulate the specific IgE response and the cellular response to allergens, treatment of patients with peanut immunotherapy is now being studied as a possible option (31). Immunotherapy with specific recombinant allergen epitopes rather than the crude allergen mixture could prove to be a more effective treatment modality. Another use of immunotherapy could be the modification of

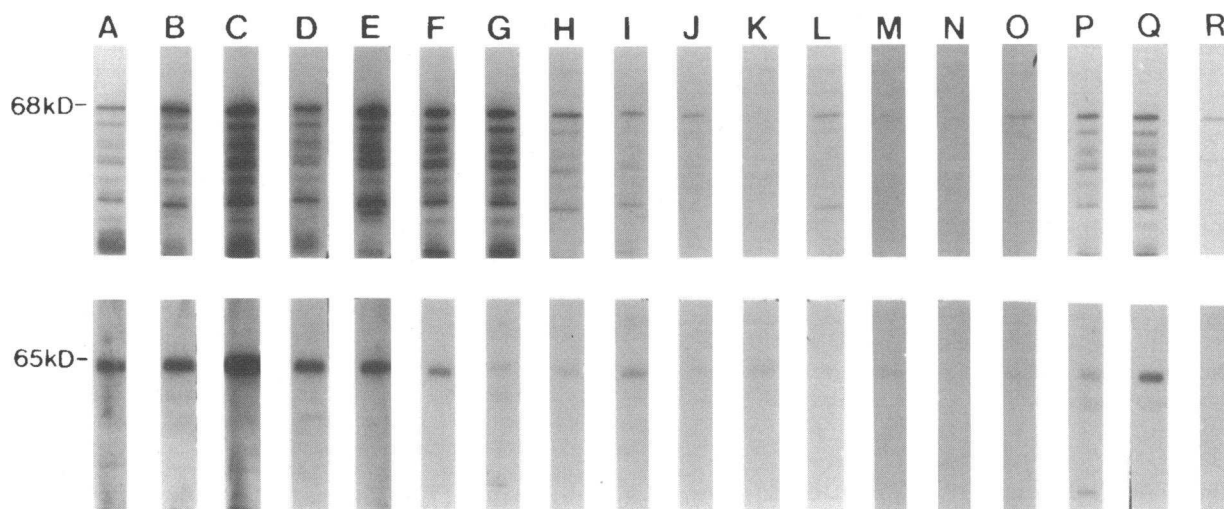


Figure 4. Serum IgE from individual patients with peanut hypersensitivity recognize recombinant and native *Ara h I* protein in an immunoblot assay. (Upper panel) *E. coli* XL1-Blue cells carrying clone P172C were induced to express the recombinant *Ara h I* protein and lysates were prepared for immunoblot analysis. Each lane (A-R) represents a different patient with peanut hypersensitivity. (Lower panel) Lanes A-R, purified, native *Ara h I* protein was used in an immunoblot assay with serum IgE from the same individuals in the Upper Panel.

the molecular structure of the recombinant allergen in order to reduce the IgE binding capacity while retaining the T cell reactivity or the production of specific T cell epitopes designed for immunotherapy. The use of recombinant allergens in standard allergen immunotherapy would have several advantages over natural allergens, including better control of the batch to batch variability of the specific allergens and the assurance of the representation of minor allergens in standard amounts. Our finding that recombinant *Ara h I* is recognized by a large proportion of people with peanut hypersensitivity will allow these immunotherapeutic approaches to be rigorously tested.

In the last several years the primary structure of a number of important inhaled allergens have been defined by cDNA cloning, including the house dust mite Group I allergen *Der p I* (32), the Birch pollen allergen *Bet v I* (33), and one of the ragweed allergens *Amb a II* (34). Similar work with foods has identified the major allergen in shrimp as a 34-kD heat-stable protein that has > 85% homology with tropomyosin for *Drosophila melanogaster* (35). Additionally, two different cDNA clones have been isolated in maize that share significant homology to several flower-expressed gene products, including short ragweed (36). Recent work by Valenta et al. (37, 38) has shown that plant profilins are prominent allergens that can be isolated from pollens of birch, grass, and weeds. There has now been isolated a wheat profilin by cDNA cloning that appears to be a major food allergen (39).

The information we have gathered about the peanut *Ara h I* allergen and the recent identification of other recombinant food allergens should allow our understanding of the pathophysiologic and immunologic mechanisms involved in food hypersensitivity reactions to move forward. Future studies with the recombinant peanut allergens will allow a better understanding of the relevant allergens in peanuts, the human immune response involved in this hypersensitivity reaction, and the possible diagnostic and therapeutic capabilities of recombinant food allergens.

Acknowledgments

The authors wish to thank Janette Breeding for her assistance in the preparation of this manuscript.

This work was supported in part by grants from the National Institutes of Health R29AI26695-05, the Asthma and Allergy Foundation of America, and the United States Department of Agriculture - AMD:9402890.

References

1. Sampson, H. A. 1983. Role of immediate food hypersensitivity in the pathogenesis of atopic dermatitis. *J. Allergy Clin. Immunol.* 71:473-480.
2. Sampson, H. A. 1985. Food hypersensitivity and atopic dermatitis: evaluation of 113 patients. *J. Pediatr.* 107:669-675.
3. Yunginger, J. W., D. L. Squillace, R. T. Jones, and R. M. Helm. 1989. Fatal anaphylactic reactions induced by peanuts. *Allergy Proc.* 10:249-253.
4. Sampson, H. A., L. Mendelson, and J. P. Rosen. 1992. Fatal and near-fatal anaphylactic reactions to food in children and adolescents. *N. Engl. J. Med.* 327:380-384.
5. Bock, S. A., and F. M. Atkins. 1989. The natural history of peanut allergy. *J. Allergy Clin. Immunol.* 83:900-904.
6. Burks, A. W., L. W. Williams, R. M. Helm, C. Connaughton, G. Cockrell, and T. J. O'Brien. 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.
7. Chapman, M. D. 1989. Purification of allergens. *Curr. Opin. Immunol.* 1:647-653.
8. Anderson, J. A., and D. D. Sogn, eds. 1984. Adverse reactions to foods. NIH Publication No. 84-2442, July, pp 1-6.
9. Metcalfe, D. D. 1985. Food allergens. *Clin. Rev. Allergy.* 3:331-349.
10. Lemanske, R. F., and S. L. Taylor. 1987. Standardized extracts, Foods. *Clin. Rev. Allergy.* 5:23-36.
11. Thomas, W. R., G. A. Stewart, R. J. Simpson, K. Y. Chua, T. M. Plozza, R. J. Dilworth, A. Nisbet, and K. J. Turner. 1987. Cloning and expression of DNA coding for the major house dust mite allergen *Der p I* in *Escherichia coli*. *Int. Arch. Allergy Appl. Immunol.* 85:127-129.
12. Knox, R. B., M. B. Singh, T. Hough, and P. Theerakulpisut. 1989. The rye grass pollen allergen *Lol p I*. In *Allergy and Molecular Biology*, Adv. Biosci. 74:161-171.
13. Burks, A. W., L. W. Williams, C. Connaughton, G. Cockrell, T. O'Brien, and R. M. Helm. 1992. Identification and characterization of a second major peanut allergen, *Ara h II*, utilizing the sera of patients with atopic dermatitis and positive peanut challenge. *J. Allergy Clin. Immunol.* 90:962-969.
14. Larsen, J. N., P. Stroman, and H. Ipsen. 1992. PCR based cloning and sequencing of isogenes encoding the tree pollen major allergen *Car b I* from *Carpinus betulus*, Hornbeam. *Mol. Immunol.* 29:703-711.
15. Bannon, G. A., F. J. Calzone, J. K. Bowen, C. D. Allis, and M. A. Gorovsky. 1983. Multiple, independently regulated, polyadenylated messages for histone H3 and H4 in *Tetrahymena*. *Nucleic Acids Res.* 11:3903-3917.
16. Watson, C. J., and R. F. Jackson. DNA Cloning. Vol I. D. M. Glover, editor. IRL Press. 79-88.
17. Huynh, T. Y., R. A. Young, and R. W. Davies. 1985. Constructing and

screening of cDNA libraries in lambda gt10 and lambda gt11. In DNA Cloning, A Practical Approach Vol I, D. M. Glover, editors. Oxford Press. 49-78.

18. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463-5467.

19. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (Lond.)*. 227:680-685.

20. Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedures and some applications. *Proc. Natl. Acad. Sci. USA.* 76:4350-4354.

21. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857-872.

22. Coleman, J., M. Inukai, and M. Inouye. 1985. Dual functions of the signal peptide in protein transfer across the membrane. *Cell.* 43:351-360.

23. Shatzman, A. R. and M. Rosenberg. 1987. Expression, identification, and characterization of recombinant gene products in *Escherichia coli*. *Methods Enzymol.* 152:661.

24. Wood, C. R., M. A. Boss, T. P. Patel, and J. S. Engle. 1984. The influence of messenger RNA secondary structure on expression of an immunoglobulin heavy chain in *Escherichia coli*. *Nucleic Acids Res.* 12:3937.

25. Yunginger, J. W., and R. T. Jones. 1987. A review of peanut chemistry: Implications for the standardization of peanut extracts. In Regulatory Control and Standardization of Allergenic Extracts. Proceedings of The Fourth International Paul Ehrlich Seminar, M. Schaeffer, C. Sisk, H. D. Brede, editors. Washington DC. 251-264.

26. Bush, R. K., S. L. Taylor, and J. A. Nordlee. 1989. Peanut sensitivity. *Allergy Proc.* 10:261-264.

27. Taylor, S. L., W. W. Busse, M. I. Sachs, J. L. Parker, and J. W. Yunginger. 1981. Peanut oil is not allergenic to peanut-sensitive individuals. *J. Allergy Clin. Immunol.* 68:372-375.

28. Chee, P. P., and J. L. Slightom. 1991. Molecular biology of legume vicilin-type seed storage protein genes. *Sub-Cellular Biochemistry.* 17:31-52.

29. Dure, L. 1990. An unstable domain in the vicilin genes of higher plants. *The New Biol.* 2(5):487-493.

30. Valentine, M. D., K. C. Schuberth, A. Kagey-Sobotka, D. F. Graft, K. A. Kwitrovich, M. Szilko, and L. M. Lichtenstein. 1990. The value of immunother-

apy with venom in children with allergy to insect stings. *N. Engl. J. Med.* 323:1601-1604.

31. Oppenheimer, J. J., H. S. Nelson, S. A. Bock, F. Christensen, and D. Y. M. Leung. 1992. Treatment of peanut allergy with rush immunotherapy. *J. Allergy Clin. Immunol.* 90:256-262.

32. Chua, K. Y., G. A. Stewart, W. R. Thomas, R. J. Simpson, R. J. Dilworth, T. M. Plozza, and K. J. Turner. 1988. Sequence analysis of cDNA coding for a major house dust mite allergen, *Der p 1*. Homology with cysteine proteases. *J. Exp. Med.* 167:175-182.

33. Breiteneder, H., K. Pettenburger, A. Bito, R. Valenta, D. Kraft, H. Rumpold, O. Scheiner, and M. Breitenbach. 1989. The gene coding for the major birch pollen allergen *Bet v 1*, is highly homologous to a pea disease resistance response gene. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:1935-1938.

34. Rogers, B. L., J. P. Morgenstern, I. J. Griffith, X. B. Yu, C. M. Counsell, A. W. Brauer, T. P. King, R. D. Garman, and M. C. Kuo. 1991. Complete sequence of the allergen *Amb a 11*. Recombinant expression and reactivity with T cells from ragweed allergic patients. *J. Immunol.* 147:2547-2552.

35. Shanti, K. N., B. M. Nartin, S. Nagpal, D. D. Metcalf, and P. V. Rao. 1993. Identification of tropomyosin as the major shrimp allergen and characterization of its IgE-binding epitopes. *J. Immunol.* 151(10):5354-5363.

36. Turcich, M. P., D. A. Hamilton, and J. P. Mascarenhas. 1993. Isolation and characterization of pollen-specific maize genes with sequence homology to ragweed allergens and pectate lyases. *Plant Molecular Biology.* 23(5):1061-1065.

37. Valenta, R., M. Duchene, K. Pettenburger, C. H. Sillaber, P. Valent, P. Bettelheim, M. Breitenbach, H. Rumpold, D. Kraft, and O. Scheiner. 1991. Identification of profilin as a novel pollen allergen: IgE autoreactive in sensitized individuals. *Science (Wash. DC).* 253:557-560.

38. Valenta Duchene, M., C. Ebner, P. Valent, C. Sillaber, P. Deviller, F. Ferreira, M. Tejkl, H. Edlmann, D. Kraft, and O. Scheiner. 1992. Profilins constitute a novel family of functional plant pan-allergens. *J. Exp. Med.* 175:377-385.

39. Rihs, H. P., K. Rozynek, May-Taube, B. Welticke, and X. Baur. 1994. Polymerase chain reaction based cDNA cloning of wheat profilin: a potential plant allergen. *Int. Arch. Allergy Immunol.* 105:190-194.