

Identification and cloning of a complementary DNA encoding a vicilin-like proprotein, Jug r 2, from English walnut kernel (*Juglans regia*), a major food allergen

Suzanne S. Teuber, MD,^{a,b} Koren C. Jarvis, BS,^a Abhaya M. Dandekar, PhD,^c W. Rich Peterson, BS,^a and Aftab A. Ansari, PhD^d Davis and Pleasant Hill, Calif, and Atlanta, Ga

Background: Walnuts and other tree nuts are important food-allergen sources that have the potential to be associated with life-threatening, IgE-mediated systemic reactions in some individuals.

Objective: The purpose of this study was to characterize a complementary (c)DNA clone encoding one of the walnut food allergens.

Methods: A cDNA expression library prepared from walnut somatic embryo was screened for IgE reactivity with patient serum. A reactive clone of 2060 bp, which encoded a protein of 593 amino acids in length, was subcloned by excision into the pGEX expression vector. IgE-binding inhibition experiments were performed.

Results: A recombinant fusion protein was induced and shown to bind serum IgE from 9 of 15 patients tested, thus identifying a major allergen. This clone, named Jug r 2, exhibited significant homology with genes encoding the vicilin group of seed proteins. An IgE-binding inhibition experiment suggested that the encoded protein undergoes posttranslational modification into at least one major polypeptide (47 kd) and possibly several others, which is similar to the vicilin-like proteins characterized in cocoa bean (*Theobroma cacao*) and cottonseed (*Gossypium hirsutum*). N-terminal sequencing of the 47-kd band, Jug r 2, identified it as a mature protein obtained from the precursor. A second IgE-binding inhibition experiment showed that there is minimal or no cross-reactivity between Jug r 2 and pea vicilin, peanut proteins, or cacao proteins.

Conclusion: Jug r 2 is the third vicilin food allergen identified in addition to vicilins from soy and peanut. The availability of

recombinant food allergens should help advance studies on the immunopathogenesis and possible treatment of IgE-mediated food hypersensitivity. (*J Allergy Clin Immunol* 1999;131:11-20.)

Key words: Food allergy, cDNA cloning, recombinant walnut allergen, walnut allergy, vicilin, 7S globulin

IgE-mediated food hypersensitivity reactions involving tree nuts are sometimes severe and can result in fatality.¹ Allergies to milk, egg, wheat, and soy in infants are usually outgrown by age 3 years, but allergies to fish, shellfish, peanut, and tree nuts are usually permanent and often life-threatening, with these allergens accounting for the majority of food hypersensitivity reactions in adults.² Although the precise prevalence of life-threatening tree nut allergy is unknown, a survey suggests that 0.5% of the US population is afflicted by various degrees of nut allergy.³ A survey of the clinical records at our institution has shown that the English walnut is often implicated as the source of life-threatening nut allergy. Walnuts rank third in per capita consumption of tree nuts in the United States and can be found in many foods.⁴ Little is known about the allergens involved in tree nut sensitivity, but clinically, patients often report sensitivity to more than one nut. We have recently described Jug r 1, a low-molecular-weight, 2S-albumin, seed-storage protein in walnut kernels as a major allergen.⁵ Additionally, the major allergen in Brazil nut, Ber e 1, was also described as a 2S-albumin storage protein that exhibits significant sequence homology to Jug r 1.^{5,6}

Seed proteins have been characterized on the basis of their solubility and sedimentation properties throughout the history of their study.⁷ Albumin proteins are soluble in water and low salt concentrations, whereas globulins are soluble in high salt concentrations. Legume terminology has been used by some plant geneticists to group the globulins of higher plants.⁸ The 7S globulins from legumes are referred to as vicilins and are quantitatively important seed storage proteins, which often exist as large trimeric oligomers in the seed protein bodies, with individual subunits usually 40 to 70 kd in size. In some species there appears to be a single gene encoding the subunits, and in others there appears to be a family of genes, as in soybean.⁹ Vicilin proteins have been given other names in the course of their study: conarachin of

From ^athe Division of Rheumatology, Allergy and Clinical Immunology, Department of Internal Medicine, University of California, Davis, School of Medicine, Davis; ^bVeterans Affairs Northern California Health Care System, Pleasant Hill; ^cthe Department of Pomology, University of California, Davis, Davis; and ^dthe Department of Pathology, Emory University School of Medicine, Atlanta.

S. S. Teuber was supported in part by National Institutes of Health Clinical Nutrition Research Unit grant no. DK 35747 and in part by a grant from the Allergy and Immunology Institute of the International Life Sciences Institute Foundation.

The opinions expressed herein are those of the author and do not necessarily represent the views of ILIS.

Received for publication Dec 17, 1998; revised Aug 20, 1999; accepted for publication Aug 20, 1999.

Reprint requests: Suzanne S. Teuber, MD, Division of Rheumatology, Allergy and Clinical Immunology, University of California, Davis, School of Medicine, One Shields Ave, TB 192, Davis, CA 95616.

1/1/102442

Abbreviations used

GST: Glutathione-S-transferase
PDH: Pyruvate dehydrogenase

peanut (Ara h 1 from *Arachis hypogaea*), phaseolin of common bean (*Phaseolus vulgaris*), β -conglycinin of soybean (*Glycine max*), and canavalin of jack bean (*Canavalis ensiformis*). Globulin proteins from several nonlegume plant seeds have been found to have significant sequence homology to the vicilins and can thus be designated as part of the vicilin group of proteins or vicilin-like proteins, such as α -globulin from cotton (*Gossypium hirsutum*).⁸ This article describes the cloning and sequencing of a new vicilin-like, seed-storage protein precursor, Jug r 2, from a walnut complementary (c)DNA library.

METHODS**Human sera**

The study reported herein was initiated after approval by the Human Subjects Review Committee at the University of California, Davis. Sera from 15 patients with clear, convincing histories of potentially life-threatening systemic allergic reactions to walnuts were used for this study. The clinical characteristics of these patients have been described.⁵ Control sera were obtained from patients with no history of walnut food allergy but who did have a history of allergic rhinoconjunctivitis and high specific IgE levels (class 4 or above as determined by Pharmacia ImmunoCAP) to multiple aeroallergens.

Construction and screening of a cDNA expression library

The cDNA library used in this experiment was constructed and screened as previously described.⁵ One of the clones, designated Jug r 1, was previously evaluated. A second clone was designated Jug r 2 by following the published guidelines of the International Union of Immunological Societies' Allergen Nomenclature Subcommittee¹⁰ and is the subject of the studies reported herein.

Excision and subcloning of cDNA

The recombinant Bluescript phagemids containing the cDNA of interest were excised from the λ Uni-Zap phage by using the ExAssist helper phage (Stratagene) and plated onto *Escherichia coli* strain SOLR (Stratagene) cells. The SOLR cells containing the plasmid with the cDNA insert were grown in Luria-Bertani ampicillin broth overnight at 37°C. The plasmid DNA was isolated by using the QIAprep Spin Plasmid Miniprep Kit (Qiagen) and submitted for sequencing (see below). The insert was isolated by using a partial digest because of an internal *Xho*I site in the cDNA insert. The 2-kb *Eco*R1-*Xho*I fragment was purified from a 0.9% agarose gel by using the QIAquick Gel Extraction Kit (Qiagen). The purified insert was subcloned in the correct reading frame into the pGEX-4T-3 plasmid (Pharmacia Biotech) and transformed into protease-deficient *E coli* strain BL21 (Pharmacia). The plasmid-cDNA transition site was sequenced to ensure that the cDNA was inserted in the right orientation and in frame. The pGEX plasmid expression vector results in the synthesis of foreign proteins as a fusion product with a 26-kd glutathione-S-transferase (GST).

Nucleotide sequence determination of cDNA

The plasmid cDNA was isolated and purified from transformed *E coli* strain SOLR (Stratagene) by miniprep. The cDNA underwent dideoxy sequencing with Sequenase (United States Biochemical Corp) from double-stranded DNA by using dye, labeled primers, and terminators.¹¹ Gels were run on an Applied Biosystems 373A sequencer (Applied Biosystems, Inc). Nucleotide and protein databases were searched by using the BLAST 2.0 program accessible at the Web site of the National Center for Biotechnology Information (National Library of Medicine, National Institutes of Health: www.ncbi.nlm.nih.gov).¹² Deduced amino acid sequences were aligned by using MacDNAsis software (Higgins-Sharp program, Hitachi).

Expression and purification of recombinant Jug r 2 fusion protein

Because of the partial insolubility of the fusion protein, the induction and purification of the recombinant (r)Jug r 2 fusion protein was based on the protocol described by Frangioni and Neel¹³ using *E coli* strain BL21. The fusion protein was then isolated from the cell supernatant after lysis by using the Bulk GST Purification Module (Pharmacia Biotech). The protein concentration was determined by using the Bio-Rad Protein Assay Kit (Bio-Rad). The reactivity and specificity of the fusion protein was checked with patient and control sera; approximately 1 μ g of fusion protein was loaded per lane for immunoblotting, as described below. Recombinant human pyruvate dehydrogenase (rPDH), an irrelevant GST fusion protein purified in a similar manner, was likewise loaded at 1 μ g/lane as a control.

Western immunoblotting with patient sera

SDS-PAGE, transfer to nitrocellulose membranes, and immunoblotting were performed as previously described, with the modifications that 12% acrylamide gels were also used and sera was diluted 1:10 for general use.⁵

Western immunoblotting of *E coli* lysate containing rJug r 2 by using anti-GST

SDS-PAGE, transfer, and blocking were carried out as described previously.⁵ Twenty microliters per lane of a 1:100 dilution of each of the following cell lysates was used: rJug r 2 in *E coli* strain BL21 (Pharmacia), rJug r 2 in *E coli* strain DH5 α (Promega), pGEX-4T-3 alone in strain BL21 (ie, GST without fusion protein), or *E coli* strain BL21 alone. Strips were incubated for 60 minutes at room temperature with a 1:1000 dilution of goat anti-GST antibody (Pharmacia Biotech), washed as above, and incubated for 60 minutes at room temperature with a 1:500 dilution of rabbit anti-goat IgG conjugated to alkaline phosphatase (Sigma). The strips were washed and then developed with nitroblue-tetrazolium (Sigma), at 330 μ g/mL, and 5-bromo-4-chloro-3-indolyl phosphate-p-toluidine salt (Sigma) at 164 μ g/mL in a diluent of 100 mmol/L TRIS (pH 9.5), 100 mmol/L NaCl, and 5 mmol/L MgCl₂. The development was stopped by washing the strips in 1 mmol/L EDTA.

Protein extracts

Juglans regia (cultivar Chandler) nuts were obtained from the University of California, Davis, Pomology Department, and peanuts (*Arachis hypogaea*, Spanish) were purchased raw from a local store; protein extracts were made as previously described.⁵ Purified pea vicilin was a gift from the laboratory of Shridhar K. Sathe (Florida State University, Tallahassee, Fla). Perennial ryegrass (*Lolium perenne*) pollen was purchased from Bayer (Bayer Inc), and a 1:10 wt/vol extract was prepared by suspension of the pollen in PBS at

4°C overnight on a rocker followed by centrifugation. *Theobroma cacao* glycerinated extract (1:10) was purchased from Meridian Bio-medical, Inc, and was verbally confirmed by the company to be an extract from the unprocessed cacao seed.

Absorption of human sera with rJug r 2 polypeptide and immunoblotting against walnut kernel extract

Four human sera were diluted 1:5 in blocking buffer and incubated with the purified rJug r 2 fusion protein (70 µg/mL) or, for purposes of control, GST-PDH fusion protein (rPDH) overnight at room temperature. Immunoblotting against walnut kernel extract was carried out as described above.

Absorption of human sera with peanut, pea vicilin, ryegrass pollen, and cacao extracts and blotting against walnut kernel extract

Sera from the 4 patients used in the absorption experiment above were titrated (1:5, 1:10, 1:20, 1:40, 1:80, and 1:160) against walnut kernel extract to determine an appropriate further dilution to use in experiments that would favor detection of cross-reactivity with other vicilin proteins (ie, more dilution than 1:5 while still showing IgE binding clearly to Jug r 2). Sera were then preabsorbed with increasing amounts of walnut kernel protein (1, 2, 5, 10, 20, 40, 80, 160, 320, and 640 µg/mL) to determine the minimum amount needed to absorb out all IgE binding to walnut, after which sera were preabsorbed with that same amount of walnut, peanut, pea vicilin, cacao extract, and 10 times the amount of each overnight at 4°C. Immunoblotting was carried out against walnut kernel extract as above, except that 15% acrylamide was used. Preabsorption with the identical concentrations of perennial ryegrass pollen protein extract was used as a control (ie, not expected to have a vicilin-like storage protein).

N-terminal amino acid sequencing

SDS-PAGE was carried out as above on the walnut kernel extract, and the proteins were transferred to Trans-Blot PVDF membrane (0.2 µm; BioRad). The N-terminal sequencing was determined from the blot on an ABI 477A sequencer with an on-line 120A HPLC system (Applied Biosystems, Inc). Data was collected by using the ABI 610 software (Applied Biosystems, Inc).

RESULTS

Identification of an IgE-reactive clone from the *J regia* cDNA library

The clone described herein was designated Jug r 2 and was submitted to the International Union of Immunological Societies' Allergen Nomenclature Sub-Committee.¹⁰ A partial digest of the plasmid DNA by *Eco*R1 and *Xho*I produced a fragment approximately 2060 bp in length (data not shown).

Nucleotide and deduced amino acid sequence of Jug r 2

The nucleotide sequence for the 2057 base pairs and the corresponding deduced sequence of 593 amino acids from the longest open-reading frame is shown in Fig 1. The recovered fragment has an intact poly-A tail but is missing the start codon. There is a nontranslated region containing the polyadenylation consensus sequence AATAAA at position 2000. A GenBank search revealed

TABLE I. The percentage identity and similarity between the deduced amino acid sequence of the walnut vicilin-like precursor and others

Source	NCBI accession No.	% Identity	% Similarity
Cotton (<i>Gossypium hirsutum</i>) ¹⁵	137578	51	73
Cacao (<i>Theobroma cacao</i>) ¹⁶	418908	44	69
Peanut (<i>Arachis hypogaea</i>) ¹⁴	1168390	36	70
Pea (<i>Pisum sativum</i>) ¹⁷	137582	39	64
Soybean (<i>Glycine max</i>) ¹⁸	121282	37	64

NCBI, National Center for Biotechnology Information.

that the Jug r 2 cDNA codes for a vicilin-like protein that has a significant similarity score of 70% to the deduced vicilin protein in peanuts encoded by Ara h 1, the vicilin cDNA from peanuts.¹⁴ The vicilin-like precursor from cacao (*Theobroma cacao*) is also related (Table I).¹⁵⁻¹⁸ Fig 2 is a comparison of the deduced amino acid sequences of cacao, peanut, and walnut vicilin precursor proteins. The molecular weight of the newly defined pro-protein was calculated to be approximately 66 kd, which is similar to the 67-kd molecular weight previously described for the cacao precursor protein. It is likely that the missing sequence encompasses a very small segment that includes nucleotides coding for a hydrophobic signal sequence peptide. This deduction is supported by analogy to the available complete sequence of the related pea vicilin and French bean phaseolin. The short hydrophobic leader peptide in the case of the pea vicilin and French bean phaseolin is cleaved as the polypeptide is transported across the endoplasmic reticulum, leaving an initial hydrophilic region, as seen in the hydropathy plot of Jug r 2 (Fig 3).^{9,19,20} We can also deduce from Fig 3 that the mature native walnut vicilin will possibly be approximately 48 kd (not 66 kd as with Ara h 1) because the hydrophilic 170 amino acid region is cleaved in the nonlegumes cotton and cacao.^{15,16}

Expression of rJug r 2 fusion protein

The Jug r 2 cDNA was subcloned into the frame-shifted plasmid expression vector pGEX-4T-3. The fusion protein was induced, purified, and then run on SDS-PAGE and detected as a band of approximately 92 kd along with multiple truncated proteins. The GST portion of the fusion protein has a molecular weight of 26 kd, and therefore the deduced molecular weight of rJug r 2 was estimated to be 66 kd, which is in agreement with the estimate from the cDNA sequence. Immunoblotting of the fusion protein products with sera from 15 patients with a history of walnut food allergy showed that the fusion protein bound IgE from 9 (60%) of 15 patients (Fig 4). The specificity of the reactivity of our patient sera was demonstrated because none of the control atopic sera nor secondary antibody alone showed detectable IgE binding to rJug r 2. In addition, no evidence of binding of serum IgE to the irrelevant recombinant control fusion protein (rPDH) was noted, suggesting specificity for the

```

1  GACGTGGCAGAGACGACGATGACGAGGAGAATCCGCGTGACCCACGAGAACAGTATCGCC 60
   R G R D D D D E E N P R D P R E Q Y R Q

61  AATGCCAGGAGTACTGCCGAGGCAAGGGCAAGGACAGCGTCAACAGCAGCAATGCCAAA 120
   C O E Y C R R Q G Q G Q R Q Q Q Q C O I

121 TCAGGTGCGAGGAGCGGTTAGAAGAGGACCAAGAAGTCAAGAAGAAAGGAGCGCCGAC 180
   R C E E R L E E D Q R S Q E E R E R R R

181 GCGGCAGAGACGTAGACGACCAGAATCCGCGAGACCCAGAACAGCGGTATGAGCAGTGTC 240
   G R D V D D Q N P R D P E Q R Y E Q C Q

241 AGCAACAGTGTGAAAGGCAACGAAGGGGTGAGGAGCAAACTATGCCGAGAGAAGGTGTG 300
   Q Q C E R Q R R G Q E Q T L C R R R C E

301 AGCAACGGAGACAACAAGAAGAAAGAGAGCGCCAACTGCGCAGAGACCGACAAGATCCCC 360
   Q R R Q Q E E R E R Q R G R D R Q D P Q

361 AACCAACAGTACCACAGATGCCAACGCCGATGCCAGATACAAGAGCAAAGCCCGGAAAGGC 420
   Q Q Y H R C O R R C Q I Q E Q S P E R Q

421 AAAGGCAATGCCAGCAACGATGCGAACGACAATACAAGGAGCAGCAAGGCAGAGAACGGG 480
   R Q C O O R C E R Q Y K E Q Q G R E R G
                               ↓
481 GACCAGAGGCAAGCCCGCGCAGAGAAAGCAGAGGCCGGGAAGAGGAGCAACAGAGGCACA 540
   P E A S P R R E S R G R E E E Q Q R H N

541 ATCCCTACTACTTTCACTCCCAGAGCATTAGGTGAGACATGAGTCCGAGGAAGGTGAAG 600
   P Y Y F H S Q S I R S R H E S E E G E V

601 TTAAATATCTTGAGAGGTTACGGAGAGAACCAGAGCTCCTTCGTGGCATTGAGAATTACC 660
   K Y L E R F T E R T E L L R G I E N Y R
                               ♦
661 GTGTGGTTATCTTGATGCAAATCCTAACCTCCATGCTCCACATCACAAAGATGCGG 720
   V V I L D A N P N T S M L P H H K D A E

721 AATCTGTTGCCGTAGTCACAAGAGGAAGGGCGACCCCTCACCTTAGTATCTCAAGAGACAA 780
   S V A V V T R G R A T L T L V S Q E T R

781 GAGAGTCGTTCAACCTCGAGTGTGGAGATGTTATAAGGGTTCCTGCGGGTGCTACCGTGT 840
   E S F N L E C G D V I R V P A G A T V Y

841 ATGTGATTAACCAAGACAGTAACGAGAGGCTAGAGATGGTGAACTCCTCCAGCCCGTCA 900
   V I N Q D S N E R L E M V K L L Q P V N

901 ACAATCCTGGCCAATTTAGGGAATACTATGCAGCTGGTGCTAAAAGTCCTGATCAATCAT 960
   N P G Q F R E Y Y A A G A K S P D Q S Y

961 ATCTCAGGGTATTTCAGCAATGATATTCTCGTCGAGCTCTAAATACCCCGAGGGACAGGT 1020
   L R V F S N D I L V A A L N T P R D R L

1021 TAGAGAGGTTCTTCGACCAGCAGGAGCAGCGGGAAGGAGTAATCATCAGAGCCTCACAAG 1080
   E R F F D Q Q E Q R E G V I I R A S Q E

```

FIG 1. Nucleotide and deduced amino acid sequence for Jug r 2. The largest open-reading frame, beginning at position 1, encodes a polypeptide 593 amino acids in length. The cleavage site for processing of the 47-kd IgE-binding allergen Jug r 2 (44 kd on 12% acrylamide) is indicated by an *arrow* at amino acid position 173. The C-X-X-X-C motifs are *underlined*. The single potential N-linked glycosylation site at position 230 is indicated by a *diamond* above the sequence N-T-S (GenBank accession number: AF066055).

reactivity of patient IgE with epitopes present on the rJug r 2 part of the fusion protein but not present on GST. Immunoblots showing a lack of reactivity against rPDH (and hence GST) of sera from 4 patients with a high spectrum of IgE reactivity to rJug r 2 are shown on the left side of Fig 4.

The majority of fusion protein expressed was smaller than 55 kd, which, combined with likely differences in serum IgE titer to Jug r 2, may be the reason why only 4 of the patient sera demonstrated IgE binding to the small quantities of the full-length fusion protein. Absence of certain conformational epitopes on the full-length recom-

```
1081 AGAAGCTCAGGGCATTGAGCCAACACGCCATGTCTGCCGCCAGAGGCCATGGGGCCGCC 1140
      K L R A L S Q H A M S A G Q R P W G R R
1141 GCTCCTCTGGGGTCTTATCAGTCTAAAAAGCGAAAGTCCGTCTCTACTCGAATCAGTTTG 1200
      S S G G P I S L K S E S P S Y S N Q F G
1201 GGCAGTTCTTCGAGGCTTGCCCCGAGGAGCACAGGCAACTGCAAGAGATGGATGTCCTTG 1260
      Q F F E A C P E E H R Q L Q E M D V L V
1261 TCAATTACGCCGAAATCAAACGAGGAGCTATGATGGTGGCCCACTACAATTCTAAAGCCA 1320
      N Y A E I K R G A M M V P H Y N S K A T
1321 CAGTTGTAGTGTATGTGGTAGAAGGGACTGGACGCTATGAAATGCGCTGCCACATGTTT 1380
      V V V Y V V E G T G R Y E M A C P H V S
1381 CAAGCCAGAGCTACGAAGCCAAGGGAGAAGAGAACAAGAGGAGGAAGAGAGCACTGGGC 1440
      S Q S Y E G Q G R R E Q E E E E S T G R
1441 GATTCAGAAGGTAACGGCTCGTTTGGCACGTGGAGATATATTTGTAATCCCGGCAGGGC 1500
      F Q K V T A R L A R G D I F V I P A G H
1501 ATCCGATTGCGATAACTGCCTCTCAGAATGAGAATCTGCGCTTACTCGGCTTCGACATTA 1560
      P I A I T A S Q N E N L R L L G F D I N
1561 ATGGCGAGAACAAACCAGAGAGACTTCCTTGCCGGGCAAAACAACATAATTAACCAGTTAG 1620
      G E N N Q R D F L A G Q N N I I N Q L E
1621 AGAGAGAAGCAAAGGAGCTGTCGTTCACATGCCACGAGAAGAGATAGAAGAGATATTTG 1680
      R E A K E L S F N M P R E E I E E I F E
1681 AAAGCCAGATGGAATCTTATTTTGTGCCCACGGAAGGCAGAGCCGACGTGGCCAAGGGA 1740
      S Q M E S Y F V P T E R Q S R R G Q G R
1741 GAGACCACCCATTGGCATCAATTTTGGACTTCGCTTCTTCTAAGTAGGCTAGCAAAGTA 1800
      D H P L A S I L D F A F F
1801 TGTTCCAAATTATCCATATCATGCAATGCTTTGTTGGTTTGTATTGTAATAATGTCTTA 1860
1861 ATAATATAAGATAGTACTAGATCAGACGGATGAGTACTGAGGCGCTGTGTAAAAGTTTG 1920
1921 TAAGCAGCCTTATGGAGTACTTAAGTACTACTTCTGTATTTCGGATGAAGTTGTCTAGCTT 1980
1981 TTGATGATCAGTATGCATAAAATAAAACAAAGTTCCATGTGTTTGCTCAAAAAAAAAAAAA 2040
2041 AAAAAAAAAAAAAAAAAA 2057
```

FIG 1. CONT

binant protein and fragments could also be a reason for the variation in IgE binding. Truncation is a known shortcoming of bacterial expression systems, in which, for instance, rare arginine codons have been shown to be the cause of truncated expression of a related vicilin, Ara h 1.²¹ To verify that the multiple low-molecular-weight polypeptides binding IgE were indeed truncated products, immunoblotting with anti-GST as the primary antibody instead of sera was performed (Fig 5). It showed a small amount of full-length fusion protein and a large quantity of low-molecular-weight proteins. These results support the notion that the fusion protein is truncated at the time of translation rather than undergoing nonspecific degradation in the bacterial cell or in solution.

Determination of the N-terminal cleavage site of the proprotein Jug r 2

Absorption experiments were performed to determine the molecular weight of the mature vicilin-like protein to be targeted for amino acid sequencing because Jug r 2 encodes a proprotein, which, by analogy with cacao and cotton, is likely cleaved just after the highly hydrophilic sequence containing C-X-X-X-C repeats, which are characteristic of vicilin proteins, to give a protein of about 48 kd. Preabsorption of sera from 4 patients with rJug r 2 abolished IgE binding to a walnut kernel protein at approximately 44 kd in all 4 sera (approximately 47 kd on 15% acrylamide, data not shown; Fig 6). Also, IgE

		10	20	30	40	50	
Ara h1	1	MRGRVSPLML	LLGILVLASV	SATQAKSPYR	KTENPCAQR	LOSQQQEPDD	50
Cacao	1	-----	-----VIS	-----KSPFI	VLIFS---LL	LSF-----	50
Jug r2	1	R-----	-----GRD	DDDEENPRDP	REQYR---QC	QEYQRRQGQG	50
		60	70	80	90	100	
Ara h1	51	LKQKA-CESR	CT-KLEMDPR	CVYDTGATNQ	RHPPGERTRG	RQPGDYODDR	100
Cacao	51	-----ALL	CSG-VSAYGR	KOYERDPR--	-----	--QYEQCO	100
Jug r2	51	QRQQQCQIR	CEERLEEDQR	SOEERERRRG	RDVDDQNPRD	PEQ-RYEQCO	100
		110	120	130	140	150	
Ara h1	101	RPRREEGQR	WGPA-----	-----EPRE	REREEDWRQP	REDWRR----	150
Cacao	101	RCBSEATEE	RECEQCQRRC	BREYKEQQRO	QE-----EEL	QROYQQCGR	150
Jug r2	101	QOCERQRRQ	-EQTLQRRRC	QRRQQEERE	RQRGRDRQDE	QOYHRCQRR	150
		160	170	180	190	200	
Ara h1	151	-----ESH	QQP-----R	KIRPEGREGE	QEWGTGSSIV	REETS---RN	200
Cacao	151	COEQCCGORE	QOCCCKKWE	OYKEQERGEH	ENYHNHKKNR	SEEEEGQORN	200
Jug r2	151	COIQEOSPER	QROCCQRCEP	OYKEQQRGR	GPEASPRRES	RGREEEQQRH	200
		210	220	230	240	250	
Ara h1	201	NPFYFPSRR-	ESTRYGNONG	RIRVLQRFDO	RSKQFONLON	HRIIVQIFARP	250
Cacao	201	NPYYFPKRRS	EOTRFDEEG	NFKILQRFAB	NSPPLKGINP	YRLAMFANP	250
Jug r2	201	NPYYFHSQ-S	IRSRHSEEG	EVKYLRFTE	RTELRLGLEN	YRVVILLDANP	250
		260	270	280	290	300	
Ara h1	251	NTLVLPKHAD	ADNLLVLOQG	QATVTVWANGN	NKRSFNLECG	HALRIFPSGFI	300
Cacao	251	NTFTLPHHCD	AEALYFVING	KGITLFIHIE	NKESYNNVORG	TWVSVPAGST	300
Jug r2	251	NTSMLPHHKD	ABSVAIVTRG	RATLTLVSQE	TRESFNLECG	DVIRVPAGAT	300
		310	320	330	340	350	
Ara h1	301	SYILNRHDND	NIRVAKISMP	VNTPGOFEDF	FPASSRD-QS	SYLOGFSRNT	350
Cacao	301	VYVVSODNOE	KLITIAVIALP	VNSPKYECF	FPACNNKPE-	SYYGAFSYEV	350
Jug r2	301	VYVINODSNE	REEMVKILQP	VNNPGOFREY	YAAGAKSDQ	SYLRVFSNDI	350
		360	370	380	390	400	
Ara h1	351	LEAAFNAEFN	EIRRVLEEN	AGGEQEERGO	RRRSTRSSDN	EGVIVKYSKE	400
Cacao	351	LEIVENTORE	KLEEIL----	-----EEOGGO	KRQOQO----	QGMFRKAYPE	400
Jug r2	351	LVAALNTPRD	RILERFF----	-----DQO----	-----EQR----	EGVITIRASQE	400
		410	420	430	440	450	
Ara h1	401	HVOELIKHAK	SVSKKGSEEE	DITNPINLRD	GEEDLSNNFG	RLFVVKDQKK	450
Cacao	401	QIRALSOQAT	SPRHRC-GER	LA---INLLS	QSEVYSNONG	RFFFACPEP-	450
Jug r2	401	KLRLALSOHAM	SAGQRFWGRR	SSGGPISLKS	ESSEYSNOFG	QFFFACPEE-	450
		460	470	480	490	500	
Ara h1	451	NPQLQDLDM	LTCVEIHGCA	LMLPHFNSKA	MVLVVVNKCT	GNLELVAVRK	500
Cacao	451	FSQFQNDVA	VSAFKLNQGA	IFVPHYNSKA	TFVVFVIDGY	GXAQMCPHVL	500
Jug r2	451	HRQLQEMDVL	VNYAEIKRGA	MMVPHYNSKA	TVVVVYVEGT	GRYEMACPHV	500
		510	520	530	540	550	
Ara h1	501	EQQQRGRRE-	-QWELEED	EEEEGSNREV	RRYTARLKEG	DVFIMPAHFP	550
Cacao	501	SRQSGSGSG	RODRREQEEE	SEEBIFGE-F	QOVKAPLSPC	DVFVAPAGHA	550
Jug r2	501	SSQSYEGQG-	---RREQEE	--EESTGR-F	QKVTARLARG	DVFVVPAGHP	550
		560	570	580	590	600	
Ara h1	551	VAINASSE--	LHLGFGINA	ENNRIFLAG	DKONVLDQIE	KQAKDLAFPG	600
Cacao	551	VITFASKDQP	LNAVAFGLNA	QNNORIFLAG	-KQNLVRQMD	SEAKELSFGV	600
Jug r2	551	IATPASQNN	LRLLGFDING	ENNORDFLAG	-QNNIINQIE	REAKELSFNM	600
		610	620	630	640	650	
Ara h1	601	SGEOVEKLIK	NQRESHFVSA	RPOSQSPSSP	EKEDQEEINQ	GKGKPLLSIL	650
Cacao	601	PSKLVDNIFN	NPEDESIFMSF	SOQROR----	-----GDERRG	NPLASILDFA	650
Jug r2	601	PREIEEIEFE	SOMESYFVPT	EROSRR----	-----GQGE-D	HPLASILDFA	650
		660	670	680	690	700	
Ara h1	651	KAFN.....					700
Cacao	651	RLF.....					700
Jug r2	651	-FF.....					700

FIG 2. Alignment of the deduced amino acid sequence of the Jug r 2 precursor with the peanut Ara h 1 precursor (Swiss Prot accession number: P43237)¹⁴ and the cacao vicilin-like precursor (PRF accession number: 1905429A).¹⁶ Thirty-six percent identity exists between Jug r 2 and Ara h 1, and 44% identity exists between Jug r 2 and the cacao vicilin precursor. Areas of identity are shaded. The cleavage site for the mature protein is indicated by an arrow.

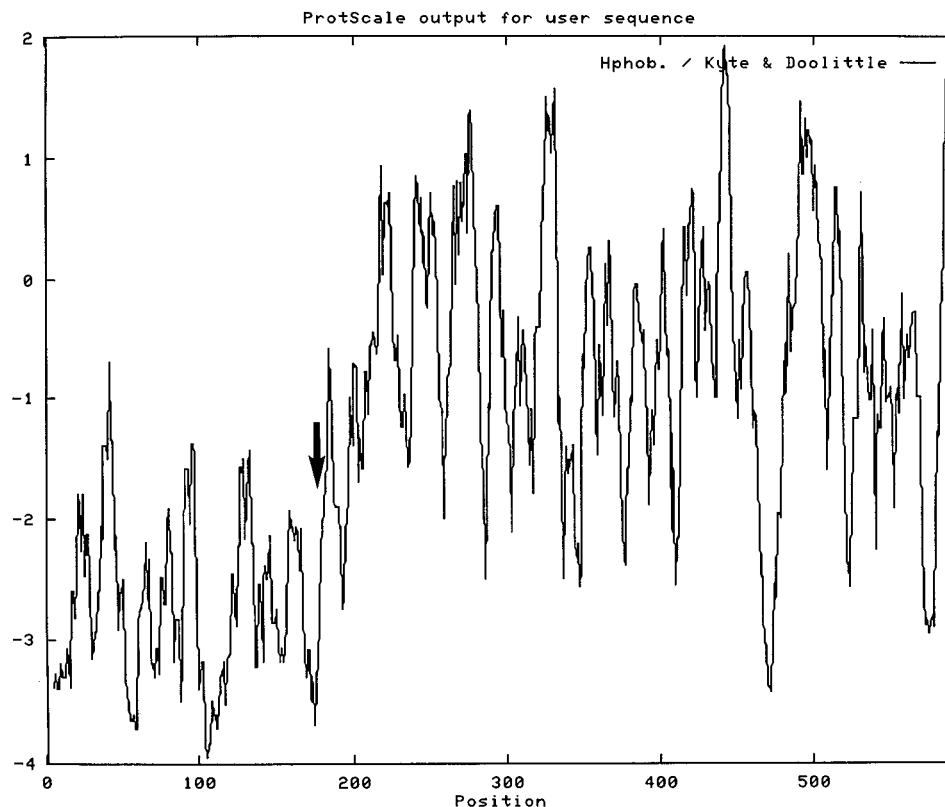


FIG 3. Hydropathy plot. Amino acid position is indicated on the horizontal axis, and the hydropathy score, as per Kyte and Doolittle,²⁰ is on the vertical axis. Positive values are hydrophobic; negative values are hydrophilic. The cleavage site for the mature protein is approximated by an arrow.

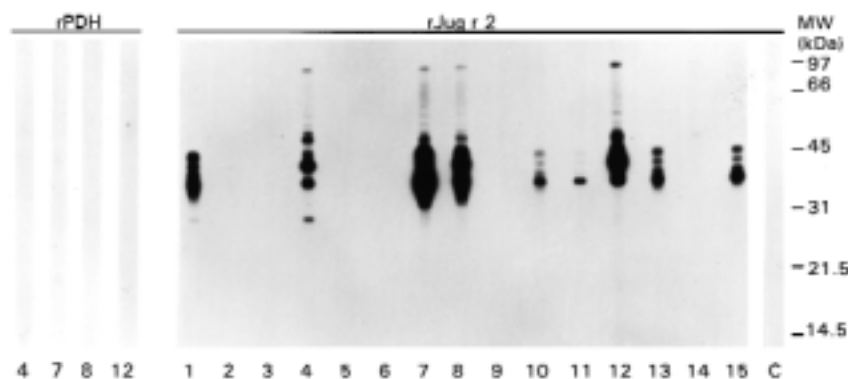


FIG 4. IgE immunoblots against recombinant proteins are shown. On the left, rPDH is probed with 4 sera (4, 7, 8, and 12), which refer to the corresponding patient sera used on the right side of the illustration. On the right (rJug r 2), lanes 1-15 refer to sera from 15 individual patients, and lane C refers to pooled control atopic sera control strip. MW, Molecular weight. rPDH is an irrelevant protein used as a control.

binding was inhibited in 3 of the patient sera to another band at approximately 48 kd on an immunoblot from a 12% acrylamide gel, suggesting that this protein also cross-reacts with rJug r 2. Some inhibition of IgE binding to proteins at about 52 and 28 kd was seen with serum from patient 4. The lack of IgE-binding inhibition with sera preabsorbed with rPDH GST-containing fusion

protein provides further evidence that all absorption was caused by specific binding of patient IgE to the recombinant walnut vicilin-like protein.

On the basis of the above, the 44-kd band (on 12% acrylamide, otherwise a 47-kd band on 15% acrylamide) was targeted for sequencing. The 44-kd band had high background for the first position (*), but thereafter a

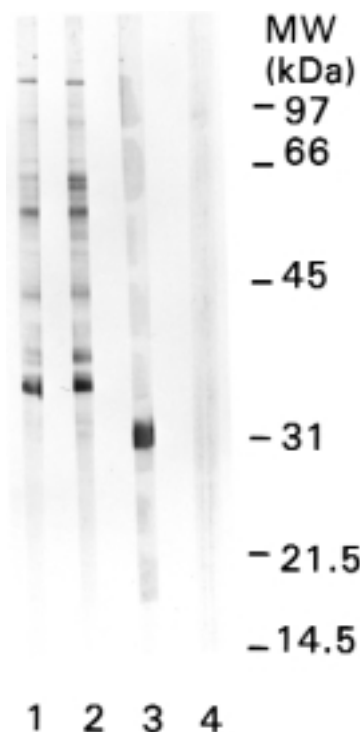


FIG 5. Anti-GST immunoblot against cell lysates. The GST protein is a single band at 26 kd. Lane 1, Jug r 2 cDNA in pGEX-4T-3 plasmid, *E coli* strain BL21; lane 2, Jug r 2 cDNA in pGEX-4T-3 plasmid, *E coli* strain DH5 α ; lane 3, pGEX-4T-3, *E coli* strain BL21; lane 4, *E coli* strain BL21 alone. MW, Molecular weight.

clean sequence was obtained: *EEEQQRHNPYYFH. From this sequence information, it was deduced that the 44-kd protein corresponds to a mature protein obtained by cleavage of the precursor at amino acid position 173; arrows in Figs 1 to 3 indicate the site. This is in concordance with the hydropathy plot (Fig 3), which predicts, on the basis of processing in cacao and cotton, that the cleavage site for the mature protein should be around amino acid 170. The estimated molecular weight of the mature protein, on the basis of this cleavage site, is 48.3 kd. To reconcile this with the estimated molecular weight of this protein from the gel of 44 kd, it is possible that there is also processing occurring at the carboxy terminus, as in the 2S-albumin seed-storage proteins,⁵ but as indicated, we have found that the 44-kd band appears as a 47-kd protein on a gel with a higher percentage acrylamide content (Fig 7). In any case this sequenced band can, with certainty, be called Jug r 2, whereas the others to which IgE from patient sera is partially absorbed out by rJug r 2 await further studies.

The fact that binding of sera IgE to more than one protein band was partially absorbed out by rJug r 2 suggests that the native vicilin-like protein may consist of several subunits derived from a single precursor molecule or perhaps that different degrees of glycosylation occur in vivo. It is also possible that multiple isoforms exist that are encoded by different genes or that related gene super-

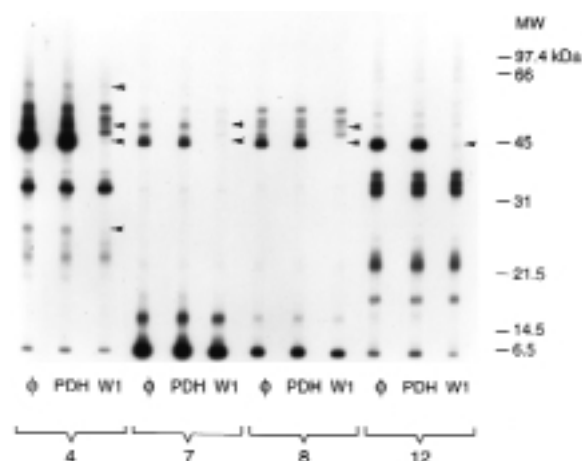


FIG 6. IgE immunoblots against walnut kernel extract separated on 12% acrylamide, reducing conditions, before (ϕ) and after absorption of the sera with an irrelevant fusion protein rPDH (PDH), or the recombinant of interest, rJug r 2 (W1). Sera from 4 patients (4, 7, 8, and 12 from Fig 4) were used. Arrows indicate the bands that were attenuated or absorbed out by the vicilin-like recombinant. MW, Molecular weight.

family products (eg, sucrose-binding proteins) have cross-reactive IgE epitopes recognized by some patients. There is precedence in other species both for the processing of the precursor protein into subunits of different sizes, as well as for the existence of a vicilin gene family encoding several different vicilin isoform precursors.^{8,9,15} Further studies beyond the scope of the data presented here are needed to delineate the posttranslational processing and the possible presence of isoforms encoded by related vicilin genes.

Cross-reactivity with other plant seed vicilins

Relatively small amounts of walnut kernel extract (1, 2, and 5 μ g/mL) were able to completely absorb out most, if not all, walnut-specific IgE in the sera from patients 4, 8, and 7, respectively (Fig 7). Sera from patient 12, however, still had faint IgE binding after absorption at 20 μ g/mL (to protein bands at 22 and 36-38 kd, which do not reproduce well in Fig 7), which was not decreased further by increasing the concentration to 640 μ g/mL. Therefore 20 μ g/mL was used as the first absorption. The IgE-binding patterns are different than those shown in Fig 6 because of differences in sera titer (more diluted), longer radiographic film exposure time, and percentage of acrylamide used for SDS-PAGE, which alters protein migration (15% here compared with 12% in Fig 6). Of note, patients 4, 7, and 12 are also allergic to peanut, and none are clinically reactive to pea or chocolate. Sera IgE from patients 4 and 8 showed no easily discernible cross-reactivity with other seed proteins in the concentrations used. IgE binding from serum of patient 7 to proteins at 6, 31, and 40 kd was partially absorbed out by preincubation with 50 μ g/mL peanut extract, but the binding of IgE to walnut Jug r 2 was basically unaffected. Somewhat similarly, in the case of patient 12, sera IgE binding to all walnut proteins except Jug r 2 was absorbed out by preincubation with

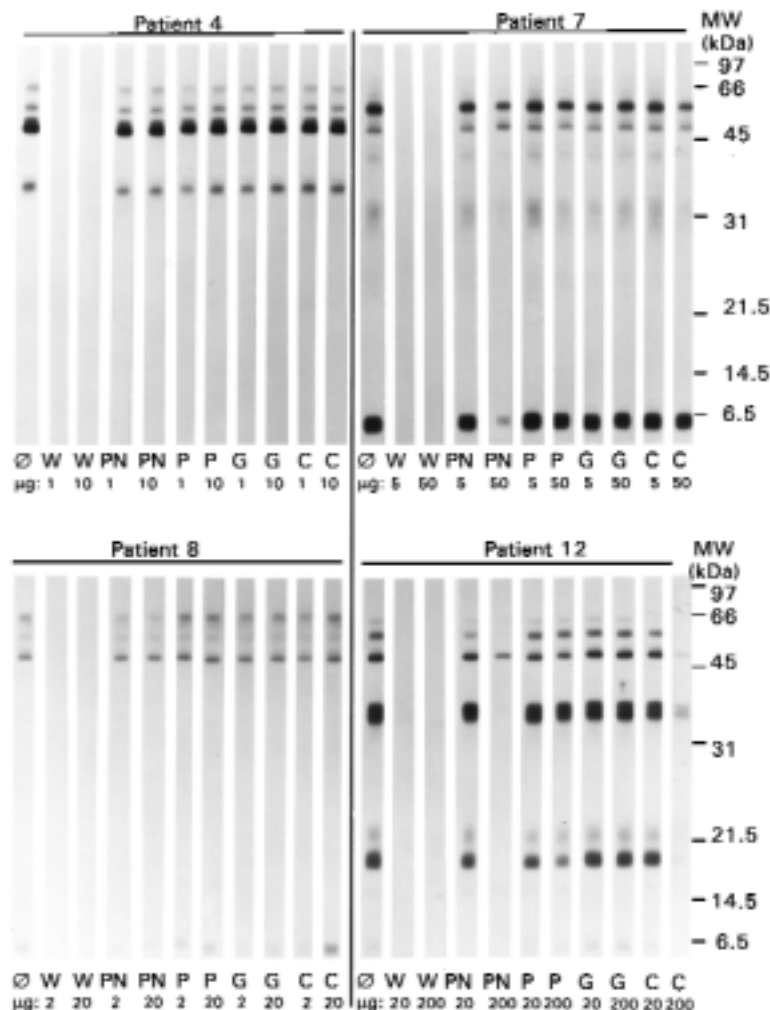


FIG 7. IgE immunoblots against walnut kernel extract run on 15% acrylamide, reducing conditions, before (Ø) and after absorption of the sera with walnut kernel extract (W), peanut kernel extract (PN), pea vicilin (P), perennial ryegrass pollen extract (G), or cacao seed extract (C) with the same 4 patient sera (4, 7, 8, and 12) used in Fig 5, but sera were diluted differently. On the bottom of each series of strips is the concentration in micrograms per milliliter of protein extract used in the absorption.

200 µg/mL peanut extract. Preabsorption of sera with the highest concentration of cacao protein extract globally inhibited binding. Overall, however, these results suggest that there is little high-affinity cross-reactivity between these seed-storage proteins because preincubation of sera with the minimal concentration of protein that had completely absorbed out IgE binding to walnut had no effect on the IgE-binding pattern to walnut when other protein sources were used. Because of the frequent concurrence of multiple tree nut and peanut allergies, future detailed immunoblot inhibition studies with more patient sera are indicated.

DISCUSSION

Vicilins in legumes are important seed-storage proteins that not only play a role as nitrogen reserves for the germinating seedling but also appear to serve a dual role

as plant defense-related proteins. In cowpea, (*Vigna unguiculata*), vicilin storage proteins have been demonstrated to inhibit fungal growth and may do so by binding to chitin, a homopolysaccharide of *N*-acetyl-D-glucosamine, in the cell wall of phytopathogenic fungi.²² In addition, cowpea vicilin binds to chitin-containing structures in the midgut of insects, and variations in the vicilin structure may influence the resistance or susceptibility of the seed to the cowpea weevil (*Callosobruchus maculatus*).^{23,24}

The vicilins show significant heterogeneity in the processing of the mature protein. Cacao and cotton both possess a hydrophobic signal sequence followed by a markedly hydrophilic region, as seen in Fig r 2, which is cleaved in cotton and presumed to be cleaved in cacao to give an approximately 47-kd mature subunit.^{15,16} Cacao and cotton both have a higher number of the C-X-X-X-C motifs (4 and 5, respectively) in the hydrophilic region

than seen in the legume vicilins (eg, soy vicilin with 2), which usually retain the hydrophilic sequence.^{15,16} Considering that the walnut vicilin has an unprecedented 6 C-X-X-X-C motif, our finding that this section is cleaved from the dominant mature protein seen at 47 kd (44 kd in 12% acrylamide) is consistent with known vicilin post-translational processing. French bean, soy, and cotton vicilins all occur in glycosylated and nonglycosylated forms; glycosylation patterns in walnut have not been reported, but the walnut vicilin has one potential N-linked glycosylation site at amino acid position 230 (N-T-S).^{9,15}

Jug r 2 is a major allergen because serum IgE from 9 (60%) of 15 patients bound to the recombinant protein. In the walnut protein body Jug r 2 is probably part of a complex trimeric structure like most other vicilins. Linear and conformational epitope mapping remain to be performed. None of the previously described nonlegume vicilins have been identified as allergens. In fact, chocolate allergy is vanishingly rare (proteins in processed chocolate are in an insoluble complex form²⁵), whereas the allergen in cottonseed has been shown to be a 2S albumin, with sequence homology to Jug r 1.²⁶ To date, only the vicilin proteins of peanut, soy, and now walnut, have been shown to be allergens.^{14,27}

Only one of our serum samples tested showed evidence of cross-reactive IgE between walnut and cacao proteins, and the cross-reactivity was only evident when a high protein concentration was used for the absorption studies. In the preliminary immunoblot-inhibition studies shown herein, minimal in vitro cross-reactivity does exist between peanut and walnut proteins (as judged by IgE binding inhibition with 2 of the 4 sera used), but no serum IgE exhibited cross-reactivity with purified pea vicilin.

In summary, an essential step has been accomplished by cloning 2 of the major allergens in the walnut, Jug r 1 and Jug r 2.⁵ These recombinant allergens can be used to learn more about the immunopathogenesis of food hypersensitivity. Complete molecular characterization of the walnut allergens will be necessary to research potential treatments for severe hypersensitivity, such as packaged plasmid DNA immunizations for walnut allergy and potentially allergy to other closely related tree nuts in animal models.

We thank Claudio Bandera for assistance with sequencing and Tom Kenny for his help with database analysis.

REFERENCES

1. Sampson HA, Mendelson L, Rosen JP. Fatal and near-fatal anaphylactic reactions to food in children and adolescents. *N Engl J Med* 1992;327:380-4.
2. Schreiber RA, Walker WA. Food allergy: facts and fiction. *Mayo Clin Proc* 1989;64:1381-91.
3. Sicherer SH, Muñoz-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;103:559-62.
4. Anonymous. Walnut Marketing Board Statistics and Economic Data, 1996. Sacramento (CA): Walnut Marketing Board; 1996.
5. Teuber SS, Dandekar AM, Peterson WR, Sellers CL. Cloning and sequencing of a gene encoding a 2S albumin seed storage protein precursor from English walnut (*Juglans regia*)—a major food allergen. *J Allergy Clin Immunol* 1998;101:807-14.
6. Nordlee JA, Taylor SL, Townsend JA, Thomas LA, Bush RK. Identification of a Brazil-nut allergen in transgenic soybeans. *N Engl J Med* 1996;334:688-92.
7. Osborne TB. The vegetable proteins. 2nd ed. London: Longman-Green; 1924.
8. Borroto K, Dure L. The globulin seed storage proteins of flowering plants are derived from two ancestral genes. *Plant Mol Biol* 1987;8:113-31.
9. Doyle JJ, Schuler MA, Godette WD, Zenger V, Beachy RN, Slightom JL. The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. *J Biol Chem* 1986;261:9228-38.
10. King TP, Hoffman D, Löwenstein H, Marsh DG, Platts-Mills TAE, Thomas W. Allergen nomenclature. *J Allergy Clin Immunol* 1995;96:5-14.
11. Tabor S, Richardson CC. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc Natl Acad Sci USA* 1987;84:4767-71.
12. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389-402.
13. Frangioni JV, Neel BG. Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. *Anal Biochem* 1993;210:179-87.
14. Burks AW, Cockrell G, Stanley JS, Helm RM, Bannon GA. Recombinant peanut allergen Ara h I expression and IgE binding in patients with peanut hypersensitivity. *J Clin Invest* 1995;96:1715-21.
15. Chlan CA, Pyle JB, Legocki AB, Dure L. Developmental biochemistry of cotton seed embryogenesis and germination XVIII. cDNA and amino acid sequence of members of the storage protein families. *Plant Mol Biol* 1986;7:475-89.
16. Spencer ME, Hodge R. Cloning and sequencing of a cDNA encoding the major storage proteins of *Theobroma cacao*. *Planta* 1992;186:567-76.
17. Watson MD, Lambert N, Delauney A, Yarwood JN, Croy RR, Gatehouse JA, et al. Isolation and expression of a pea vicilin cDNA in the yeast *Saccharomyces cerevisiae*. *Biochem J* 1988;251:857-64.
18. Harada JJ, Barker SJ, Goldberg RB. Soybean beta-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and post-transcriptional processes. *Plant Cell* 1989;1:415-25.
19. Lycett GW, Delauney AJ, Gatehouse JA, Gilroy J, Croy RRD, Boulter D. The vicilin gene family of pea (*Pisum sativum*): a complete cDNA coding sequence for preprovicilin. *Nucleic Acid Res* 1983;11:2367-80.
20. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982;157:105-32.
21. Becker WM. Characterization of Ara h 1 by two-dimensional electrophoresis immunoblot and recombinant techniques: new digestion experiments with peanuts imitating the gastrointestinal tract. *Int Arch Allergy Immunol* 1997;113:118-21.
22. Gomes VM, Mosqueda M-I, Blanco-Labra A, Sales MP, Fernandes KVS, Cordeiro RA, et al. Vicilin storage proteins from *Vigna unguiculata* (Legume) seeds inhibit fungal growth. *J Agric Food Chem* 1997;45:4110-5.
23. Sales MP, Gomes VM, Fernandes KVS, Xavier-Filho J. Chitin binding proteins from cowpea (*Vigna unguiculata*) seeds. *Braz J Med Biol Res* 1996;29:319-26.
24. Macedo MLR, Fernandes KVS, Sales MP, Xavier-Filho J. Purification and some properties of storage proteins (vicilins) from cowpea (*Vigna unguiculata*) seeds which are susceptible and resistant to the bruchid beetle *Callosobruchus maculatus*. *Braz J Med Biol Res* 1995;28:183-90.
25. Zak DL, Keeney PG. Changes in cocoa proteins during ripening of fruit, fermentation, and further processing of cocoa beans. *J Agric Food Chem* 1976;24:483-6.
26. Youle RJ, Huang AHC. Albumin storage proteins and allergens in cottonseed. *J Agric Food Chem* 1979;27:500-3.
27. Ogawa T, Bando N, Tsuji H, Nishikawa K, Kitamura K. Alpha-subunit of beta-conglycinin, an allergenic protein recognized by IgE antibodies of soybean-sensitive patients with atopic dermatitis. *Biosci Biotechnol Biochem* 1995;59:831-3.