

Characterization of Low Molecular Weight Allergens from English Walnut (*Juglans regia*)

Melanie L. Downs,^{*,†,§,¶} Aida Semic-Jusufagic,[‡] Angela Simpson,[‡] Joan Bartra,[⊗] Montserrat Fernandez-Rivas,[○] Neil M. Rigby,[#] Steve L. Taylor,[†] Joseph L. Baumert,[†] and E. N. Clare Mills[§]

[†]143 Food Industry Complex, Food Allergy Research and Resource Program, Department of Food Science and Technology, University of Nebraska—Lincoln, Lincoln, Nebraska 68583-0919, United States

[§]Institute of Inflammation and Repair, Manchester Academic Health Sciences Centre and Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom

[#]Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, United Kingdom

[‡]Centre for Respiratory Medicine and Allergy, Institute of Inflammation and Repair, University of Manchester and Respiratory and Allergy Clinical Research Facility, Education and Research Centre, University Hospital of South Manchester NHS Foundation Trust, Wythenshawe Hospital, Southmoor Road, Manchester M23 9LT, United Kingdom

[⊗]Allergy Unit, Service of Pneumology and Respiratory Allergy, Hospital Clínic (ICT), IDIBAPS, University of Barcelona, Calle Villarroel 170, 08036 Barcelona, Spain

[○]Allergy Department, Hospital Clínico San Carlos, IdISSC, c/Prof. Martín Lagos s/n, 28040 Madrid, Spain

Supporting Information

ABSTRACT: Although English walnut is a commonly allergenic tree nut, walnut allergens have been poorly characterized to date. The objective of this work was to characterize the natural, low molecular weight (LMW) allergens from walnut. A protocol was developed to purify LMW allergens (specifically 2S albumins) from English walnuts. In addition to 2S albumins, a series of peptides from the N-terminal region of the 7S seed storage globulin protein were also identified and characterized. These peptides comprised a four-cysteine motif (C-X-X-X-C-X_{10–12}-C-X-X-X-C) repeated throughout the 7S N-terminal region. Upon IgE immunoblotting, 3/11 and 5/11 sera from walnut-allergic subjects showed IgE reactivity to the 7S N-terminal fragments and 2S albumin, respectively. The mature 7S protein and the newly described 7S N-terminal peptides represent two distinct types of allergens. Because the proteolytic processing of 7S globulins has not been elucidated in many edible plant species, similar protein fragments may be present in other nuts and seeds.

KEYWORDS: 7S seed storage globulin, 2S seed storage albumin, food allergy, mass spectrometry, tree nut, walnut

INTRODUCTION

The analysis of walnut proteins has a long history, beginning with their characterization by Osborne and colleagues around the turn of the 20th century.^{1,2} These studies were among those forming the foundation of plant protein science, but since that time there has been relatively little interest in further characterizing walnut proteins. However, recently, walnut proteins have again become of interest due to their role in tree nut allergy, a potentially life-threatening condition affecting an estimated 0.6 and 1.14% of the populations of the United States and Canada, respectively.^{3,4} Walnut, specifically English walnut (*Juglans regia*), is one of the most commonly reported allergenic tree nuts,⁵ and allergic reactions to these nuts pose a serious health risk to affected individuals.^{6–9}

Four walnut proteins have been recognized by the International Union of Immunological Societies (IUIS) as major allergens: Jug r 1 (a 2S albumin), Jug r 2 (a 7S vicilin-like globulin), Jug r 3 (a nonspecific lipid transfer protein), and Jug r 4 (an 11S legumin-like globulin).^{10–13} As is the case with many plant food allergens, these walnut allergens are proteins that fall into a few superfamilies.¹⁴ The allergenic role of

proteins belonging to the 2S albumin group (in the prolamin superfamily) in peanut and other tree nut allergies has been well documented.^{15–19}

Proteins classified as 2S albumins are low molecular weight proteins (10–15 kDa) with a compact structure and a conserved pattern of cysteine residues.²⁰ Commonly, these proteins undergo posttranslational proteolysis, producing large and small subunits held together by disulfide bonds.²⁰ The 2S albumin from walnut has been identified as a major allergen, but relatively little information is available about the form(s) of the natural protein because much of the characterization has been conducted using recombinant proteins.^{10,21}

The high molecular weight allergens (Jug r 2 and Jug r 4), which belong to the cupin superfamily, are even less well characterized. Typically, 7S and 11S seed storage globulins undergo posttranslational modification including proteolytic

Received: September 29, 2014

Revised: November 7, 2014

Accepted: November 11, 2014



Table 1. Clinical Characteristics of Walnut-Allergic Subjects

location and subject ID	age	gender	walnut sIgE CAP (kU/L)	total IgE CAP (ku/L)	walnut SPT ^a	symptoms reported upon walnut ingestion
Madrid						
1	4	F	3.75	90	+	anaphylaxis ^b (erythema, OAS, ^c pruritus, abdominal pain, diarrhea, dyspnea)
2	10	F	1.95	314	–	anaphylaxis (conjunctivitis, urticaria, bronchospasm)
3	28	M	14.70	119	+	anaphylaxis (eyelid angioedema, laryngeal edema, OAS, urticaria, dyspnea, blurred vision, cold sweating, malaise)
Manchester						
4	22	M	36.00	480	nd ^d	anaphylaxis (OAS, urticaria, abdominal pain, vomiting)
5	29	F	2.00	1200	+	anaphylaxis (angioedema, urticaria, bronchospasm)
6	16	F	21.00	140	nd	anaphylaxis (lip and face angioedema, pharyngitis, urticaria, bronchospasm)
7	21	F	2.30	470	+	OAS
8	21	M	0.50	170	+	anaphylaxis (laryngeal edema, urticaria, dyspnea, wheezing, dizziness, light-headedness)
9	18	F	27.10	130	nd	anaphylaxis (lip and eyelid angioedema, face erythema, OAS, rhinitis, diarrhea, vomiting, dyspnea, wheezing, dizziness)
Barcelona						
10	28	F	1.91	82	+	anaphylaxis (urticaria, abdominal pain, diarrhea, hypotension, loss of consciousness)
11	23	F	12.10	218	+	anaphylaxis (face urticaria, abdominal pain, vomiting)

^aPositive skin prick test (SPT); wheal diameter at least 3 mm greater than negative control. ^bAnaphylaxis, as defined by World Allergy Organization criteria.³⁸ ^cOAS, oral allergy symptoms. ^dnd, not done.

processing of preproteins,²⁰ and the 7S globulin from certain plant species also undergoes N-glycosylation.²² The 7S (Jug r 2) and 11S (Jug r 4) globulin allergens of walnut appear to follow similar patterns of posttranslational modification, but the specific processing mechanisms for this plant are poorly characterized, and the presence of N-linked glycans on Jug r 2 has yet to be demonstrated.^{11,13}

Despite their early entrance into the study of plant proteins and their importance for food allergy, relatively few characterization studies have been conducted on walnut proteins, particularly regarding the natural form of the protein occurring in the nut. The objective of this study was to purify and characterize the natural 2S albumin allergen from English walnut. During the course of the study, however, an additional low molecular weight protein species was encountered with its own potential allergenic activity. Electrophoresis and mass spectrometry were employed to characterize these low molecular weight proteins, and their allergenic activity was assessed with a panel of walnut-allergic subjects' sera.

MATERIALS AND METHODS

Walnut-Allergic Subject Samples. A panel of either serum or plasma was assembled from individuals with positive walnut-specific IgE (ImmunoCAP; Thermo Scientific, Uppsala, Sweden) and a convincing history of food allergy to walnut (Table 1). The patient panel was identified from clinic populations in either the United Kingdom (Manchester Respiratory and Allergy Biobank; ManRAB) or Spain (Hospital Clínico San Carlos, Madrid or Hospital Clínic, Barcelona). A pool of plasma from nine atopic, non-food-allergic subjects from Manchester was used as a negative control. Written informed consent was obtained from all patients, either specifically for this study (Madrid and Barcelona) or for future research studies following donation (ManRAB). Ethical approvals for the use of sera were obtained from the appropriate committees (National Research Ethics Service, Research Ethics Committee, 10/H1010/7, ManRAB; Comité Ético de Investigación Clínica and Comité de Investigación del Hospital Clínic, 2012/7805, Barcelona; Comité Ético de Investigación Clínica del Hospital Clínico San Carlos de Madrid, C.P.-C.I 12/365-E, Madrid).

Materials and Walnut Preparation. English walnuts (*J. regia* cv. Chandler) were purchased from Gold River Orchards (Oakdale, CA,

USA) as raw light halves and pieces. The walnuts were ground (under liquid nitrogen, which was subsequently allowed to evaporate), and defatted walnut flour (DWNF) was produced by combining ground walnuts with hexane at a 1:5 (v/v) ratio and stirring for 1 h at room temperature. The lipid fraction was removed, the partially defatted flour was allowed to dry overnight, the dried flour was ground further, and the defatting process was repeated to produce the final DWNF (stored at –80 °C). Unless stated otherwise, all other chemical reagents were obtained from Sigma-Aldrich (Dorset, UK).

Purification of Low Molecular Weight Walnut Proteins.

Extraction of protein from DWNF was performed with a high-salt buffer (25 mM Tris-HCl, pH 7.5, containing 1.5 M NaCl, 3 mM sodium azide). Following hydration of 3.5% (w/v) poly(vinylpyrrolidone) in the buffer, soluble proteins were extracted from DWNF at a 1:50 (w/v) ratio for 1 h at room temperature, with constant stirring and pH adjustment to 7.5 every 30 min. The extract was centrifuged (3600g, 30 min, 15 °C), and the supernatant was filtered (Whatman no. 1 filter paper and sequential 0.45 and 0.2 µm syringe filters). Total soluble protein concentrations of the DWNF extract and selected protein fractions were determined by microwell BCA assay.

The chromatographic portion of the purification consisted of ConA, gel filtration, and anion exchange stages. The high salt concentration in buffers (1.5 M NaCl) was retained for ConA and gel filtration chromatography to maintain protein solubility. The ConA chromatography was conducted with a Pharmacia XK 16 column (GE Healthcare Life Sciences, Little Chalfont, UK), packed with Concanavalin A Sepharose conjugate and connected to a BioCAD Sprint LC system (Applied Biosystems, Life Technologies, Paisley, UK). Following column equilibration and walnut extract loading, unbound (non-glycosylated) proteins were eluted (in the high-salt buffer used for extraction), followed by bound (glycosylated) proteins (in 25 mM Tris-HCl, pH 7.5, containing 0.4 M methyl α-D-mannopyranoside, 1.5 M NaCl, and 3 mM sodium azide). The unbound, protein-containing fractions were pooled and concentrated 4-fold using a stirred-cell ultrafiltration apparatus (5 kDa nominal molecular weight limit (NMWL) membrane; Amicon, Millipore, Watford, UK). The concentrated, unbound fraction (UB pool) was subsequently used for gel filtration chromatography. The bound (glycosylated) fractions were not used further in this study, as they primarily contained higher molecular weight proteins (gel electrophoresis, data not shown).

Separation of UB pool proteins by gel filtration was conducted with a HiLoad 16/600 Superdex 200 column (GE Healthcare Life Sciences, calibrated with Bio-Rad (Hercules, CA, USA) Gel Filtration Standard

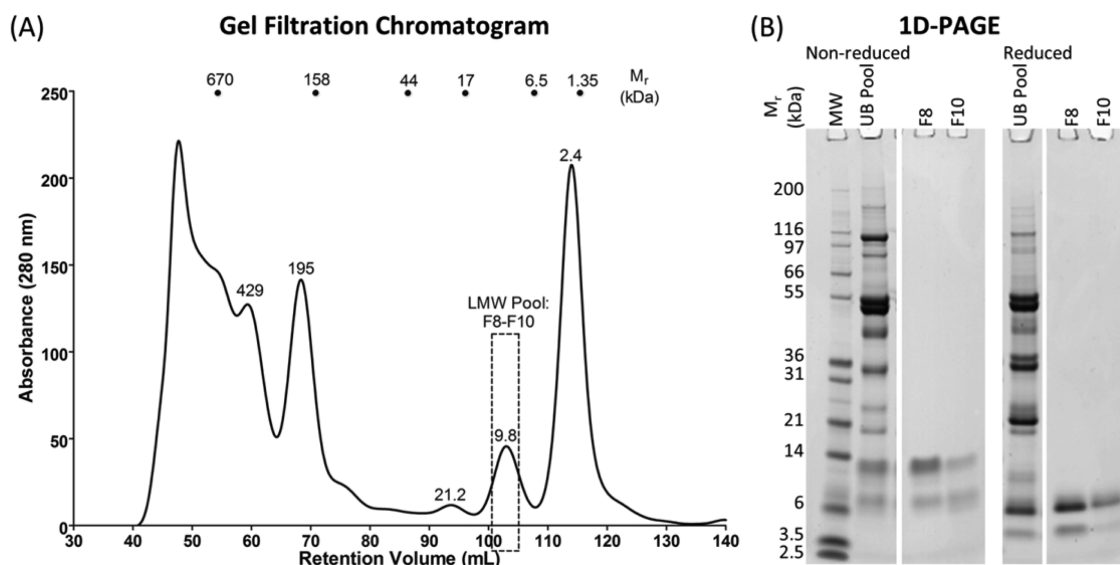


Figure 1. Gel filtration and 1D-PAGE analysis of nonglycosylated walnut proteins. (A) The gel filtration chromatogram has the calculated M_r values labeled above each major peak and a dashed box indicating the fractions included in the LMW pool. Points along the top of the chromatogram indicate retention volumes of MW standards. (B) The 1D-PAGE image includes unbound material obtained from ConA chromatography (UB pool, 7.8 μ g of protein loaded) and gel filtration fractions F8 and F10 (loaded in equal volumes) analyzed under nonreducing and reducing conditions. Molecular weight markers (MW) were also loaded (Mark12 unstained standards, Invitrogen; 5 μ L loaded).

and Sigma-Aldrich aprotinin) connected to an ÄKTA FPLC system (Amersham Pharmacia Biotech, GE Healthcare Life Sciences). Following loading, proteins were eluted with the high-salt extraction buffer over 1.2 column volumes. Fractions F8–F10 (retention volume 100.5–105 mL) were combined in equal volumes to give a pooled low molecular weight fraction (LMW pool).

The LMW pooled fraction was dialyzed (3500 Da molecular weight cutoff membrane, against 20 mM Tris-HCl, pH 8.0) and concentrated by ultrafiltration (1 kDa NMWL membrane, Millipore). Anion exchange was conducted with a POROS HQ20 column (4.6 $\times$ 100 mm) connected to a BioCAD Sprint LC system. The column was equilibrated in 20 mM Tris-HCl, pH 8.0, running buffer [40% A (50 mM Tris-HCl), 60% B (H_2O)]. The LMW pooled fraction was loaded, the column was washed (2.0 column volumes of running buffer, 3.00 mL/min), and proteins were eluted [0–300 mM NaCl linear gradient; 40% A and 60% B to 40% A, 45% B, and 15% C (2.0 M NaCl); 3.00 mL/min]. Fractions were analyzed by gel electrophoresis and mass spectrometry.

Gel Electrophoresis. One-dimensional PAGE (1D-PAGE) was conducted with NuPAGE Novex 4–12% Bis-Tris precast gels. Samples were prepared per the manufacturer's instructions, under reducing (50 mM DTT) or nonreducing conditions. Equal volumes of purification fractions (following each chromatographic stage) were loaded, along with Mark12 Unstained Standards, and separated (MES SDS running buffer, 200 V, 35 min). Gels were fixed [50% (v/v) methanol, 10% (v/v) acetic acid; 2 h], stained (SimplyBlue Safe Stain, overnight), and destained (deionized water, 2 h). Gel images were acquired with a Bio-Rad GS-800 scanner. All 1D-PAGE reagents were obtained from Invitrogen (Life Technologies).

Identification of Protein Bands by MS. Protein bands were excised from 1D-PAGE gels for in-gel trypsin digestion. Briefly, gel plugs were destained, reduced (10 mM DTT, 30 min at 60 $^{\circ}$ C), and alkylated (100 mM iodoacetamide, 30 min at room temperature in the dark) and proteins digested with trypsin (Promega Trypsin Gold, Mass Spectrometry grade, 50 ng per plug, 3 h at 37 $^{\circ}$ C). The digestion was stopped (1% (v/v) formic acid) and peptides were eluted from the gel (50% (v/v) acetonitrile). The peptide solutions were dried in a vacuum centrifuge and resuspended in 0.1% (v/v) formic acid in water for LC-MS/MS analysis.

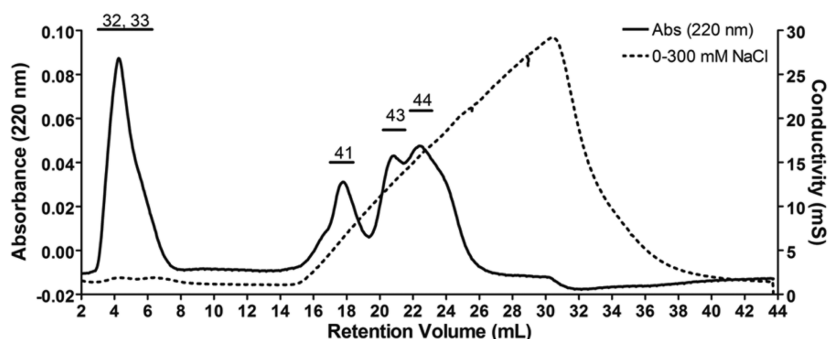
The tryptic digest samples were analyzed by LC-MS/MS (nano-ACQUITY UPLC, Waters Corp., Milford, MA; LTQ Orbitrap XL,

Thermo Scientific, Waltman, MA, USA). Peptides were trapped online to a Symmetry C18 trap column (5 μ m, 180 μ m $\times$ 20 mm; Waters Corp.), which was subsequently switched to an inline UPLC with a 75 mm $\times$ 250 μ m i.d., 1.7 μ m, BEH C18 analytical column (Waters Corp.). Peptides were separated with a gradient of 3–60% (v/v) acetonitrile in water with 0.1% (v/v) formic acid over 25 min at 250 nL min $^{-1}$. A nanospray source was used, and the mass spectrometer was operated in positive ion mode. Precursor ions were measured in the Orbitrap (60000 resolution over the mass range m/z 300–2000). The Orbitrap was operated in data-dependent acquisition mode, and the top five precursor ions (2+ or 3+ charge states) were selected for collision-induced dissociation (CID) fragmentation (dynamic exclusion enabled). Product ions were analyzed in the ion trap (rapid scan rate and normal mass range).

Mascot generic format (mgf) files were generated (Proteome Discoverer software, Thermo Scientific) and searched against the UniProt database (downloaded March 3, 2011; Viridiplantae taxonomy; 592,152 sequences with filter) using a local Mascot server (version 2.2.06, Matrix Science, London, UK). The search parameters were as follows: one missed tryptic cleavage; carbamidomethyl (C) and oxidation (M) as fixed and variable modifications, respectively; peptide mass tolerance of 5 ppm; MS/MS tolerance of 0.6 Da; and Mascot peptide significance threshold of 0.05. The MS proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository²³ (data set identifier PXD000634 and DOI 10.6019/PXD000634).

Intact Mass Determination by MS. MALDI-TOF MS was used to analyze the intact masses of the LMW proteins. The LMW pooled fraction (prior to dialysis, concentration, and anion exchange chromatography) was analyzed before and after reduction (DTT added to 5 mM, 10 min at 90 $^{\circ}$ C) and alkylation (iodoacetamide added to 10 mM, 20 min at room temperature in the dark). Samples were plated onto a polished stainless steel target with a DHB matrix [10 mg of 2,5-dihydroxybenzoic acid in 30% (v/v) acetonitrile, 70% (v/v) water with 0.1% (v/v) TFA] using the dried drop method (0.5 μ L each of matrix and sample). Calibration standards (ProteoMass Peptide & Protein MALDI-MS Calibration Kit; Sigma-Aldrich) were prepared according to the manufacturer's instructions and plated in the same manner. Intact protein spectra were acquired using a Bruker Ultraflex II MALDI-TOF/TOF instrument (Coventry, UK). Positive

(A) Anion Exchange Chromatogram



(B) 1D-PAGE

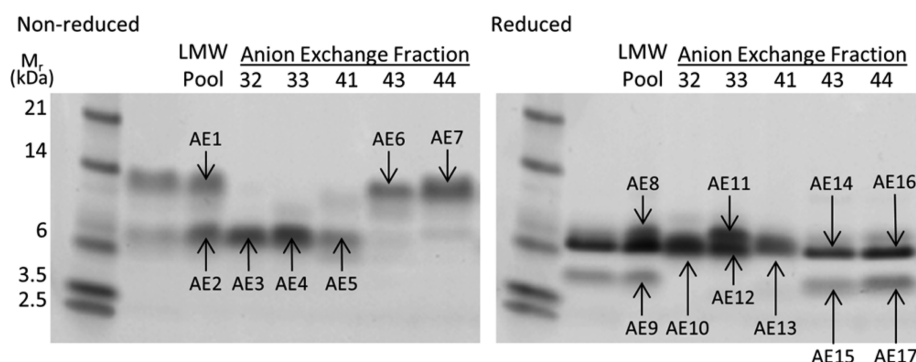


Figure 2. Anion exchange separation of LMW pool proteins and 1D-PAGE analysis of resultant fractions. (A) Anion exchange chromatogram shows separation of polypeptides in LMW pool obtained from gel filtration chromatography. (B) The 1D-PAGE of the low molecular weight (LMW) pool (1.8 μ g protein) and equal volumes of the anion exchange fractions includes analysis under nonreducing and reducing conditions, with labels indicating bands selected for LC-MS/MS analysis.

ions were analyzed in linear mode, with 200 shots per sample used to produce a summed spectrum.

IgE Immunoblotting. Proteins in the LMW pooled fraction (1.0 μ g per lane), along with SeeBlue Prestained Standards (Invitrogen), were separated by 1D-PAGE under nonreducing and reducing conditions (described above) and electroblotted onto a polyvinylidene fluoride membrane (0.2 μ m, Immobilon-P^{SO}, Millipore) using a semidry transfer apparatus (Trans-Blot SD Semi-Dry Cell, Bio-Rad). Following protein transfer, membranes were blocked with 0.01 M PBST [0.002 M NaH₂PO₄, 0.008 M Na₂HPO₄, pH 7.4, containing 0.85% (w/v) NaCl and 0.05% (v/v) Tween 20] containing 5% (w/v) nonfat dry milk (NFDm) for 2 h at room temperature with gentle shaking. Individual serum/plasma samples were diluted 1:10 (v/v) in 0.01 M PBST containing 2.5% (w/v) NFDm. Diluted serum or plasma (0.05 mL/cm²) was applied to the blocked membranes and incubated overnight at room temperature. Bound IgE was detected with a horseradish peroxidase-conjugated mouse anti-human IgE antibody (SouthernBiotech; Birmingham, AL, USA) and a chemiluminescent substrate (Pierce SuperSignal West Dura Extended Duration Substrate, Rockford, IL, USA). Images of chemiluminescence were acquired for 5 min with a CCD camera (FUJIFILM Luminescent Image Analyzer LAS-1000plus, Bedford, UK).

RESULTS AND DISCUSSION

An extract of defatted walnut flour, prepared using a high-salt buffer, was subjected to ConA affinity chromatography to selectively remove N-glycosylated proteins. Although 2S albumins are classically extracted in water, the addition of salt was found to enhance their extraction from walnut (data not shown). The resulting fraction enriched in nonglycosylated proteins was further separated using gel filtration chromatography. A protein peak with a median relative molecular mass

(M_r) of 9.8 kDa (LMW Pool, Figure 1A) was retained and analyzed by 1D-PAGE (Figure 1B). This analysis revealed the presence of two protein bands (M_r 11.8 and 6.5 kDa), which after reduction showed a change in mobility, resolving as two bands (M_r 6.0 and 4.2 kDa; Figure 1B).

The polypeptides in the LMW pool were further fractionated using anion exchange chromatography (Figure 2A). 1D-PAGE analysis of fractions from all protein-containing peaks revealed multiple protein band patterns (Figure 2B). The patterns observed with fractions 43 and 44 were analogous to the characteristic bands of other 2S albumins. Specifically, an intact protein of approximately 12 kDa under nonreducing conditions (AE6 and AE7) separated into large (6 kDa; AE14 and AE16) and small (4 kDa; AE15 and AE17) subunits upon reduction. The other fractions (32, 33, and 41) did not exhibit protein band patterns consistent with typical 2S albumin behavior. These fractions all delivered bands of 6–7 kDa on 1D-PAGE, under both nonreducing and reducing conditions (AE3, AE5, and AE10–13). The band patterns from the LMW pool demonstrated the coexistence of these two mobility patterns prior to anion exchange fractionation (AE1, AE2, AE8, and AE9).

In-gel trypsin digestion and LC-MS/MS analysis were used to identify the polypeptides observed with 1D-PAGE (Figure 2B; Table 2). A full listing of peptide spectrum matches is provided in Supporting Information Table S1. As expected, the polypeptides exhibiting typical 2S albumin mobility patterns in the LMW pooled fraction (AE1, AE8, AE9) and in fractions 43 and 44 (AE6, AE7, and AE14–17) were identified as English walnut 2S albumin (UniProt P93198; Figure 2; Table 2). This

Table 2. LC-MS/MS Identifications of Anion Exchange 1D-PAGE Bands^a

1D-PAGE band	protein (UniProt accession)	Mascot protein score
AE1	2S albumin (P93198)	366
AE2	7S globulin; pecan (B3STU4/B3STU7)	239
AE3	7S globulin; pecan (B3STU4/B3STU7)	151
AE4	7S globulin; pecan (B3STU4/B3STU7)	200
AE5	7S globulin; walnut (Q9SEW4)	224
AE6	2S albumin (P93198)	294
AE7	2S albumin (P93198)	349
AE8	2S albumin (P93198)	308
AE9	2S albumin (P93198)	177
AE10	7S globulin; pecan (B3STU4/B3STU7)	116
AE11	7S globulin; pecan (B3STU4/B3STU7)	206
AE12	2S albumin (P93198)	204
AE13	7S globulin; pecan (B3STU4/B3STU7)	226
AE14	2S albumin (P93198)	282
AE15	2S albumin (P93198)	213
AE16	2S albumin (P93198)	339
AE17	2S albumin (P93198)	186

^aThe top-ranked protein match (and UniProt accession number) for each band labeled in Figure 2B is given in the table, along with the corresponding Mascot protein score.

walnut 2S albumin was also identified as the top-ranking protein in band AE12, indicating the presence of the protein or a protein fragment in fraction 33.

The remaining low M_r polypeptides in the second set of ion exchange fractions (32, 33, 42) were identified as the 7S seed storage globulin from pecan (*Carya illinoensis*, B3STU4/B3STU7) and English walnut (Q9SEW4) (Figure 2; Table 2). This result is inconsistent with the experimentally determined M_r of the mature 7S globulin from walnut (44–48 kDa by SDS-PAGE).¹¹ Inspection of the peptide spectrum matches from these bands showed that the peptide matches were located exclusively in the N-terminal regions of the pecan and walnut 7S globulin sequences (Figure 3A). However, all peptides were not identified in all 1D-PAGE bands (Figure 3B), indicating differences among the bands or in the compatibility of the various peptides with LC-MS/MS. The localization of peptides in the N-terminal region also explains the higher protein scores for the 7S globulin from pecan, as the walnut sequence currently in the public databases is incomplete in the N-terminal region. The walnut and pecan sequences are very similar throughout the portion of the walnut 7S sequence present in the database (92% identical), and therefore the peptide spectrum matches assigned exclusively to the pecan proteins can be assumed to be homology-based matches to walnut proteins and not indicative of the presence of pecan.

As is the case with many seed storage proteins, the 7S vicilin-like globulins are frequently synthesized as a pre-proprotein, which undergoes posttranslational proteolytic processing to produce the mature protein.²⁰ In some plant species (e.g., peanut, soy, pumpkin, cocoa, and cotton) evidence indicates that along with the proteolytic removal of the signal peptide, an additional hydrophilic portion of the proprotein N-terminal region is removed during the production of the 7S

globulin.^{24–31} On the basis of N-terminal sequencing data from the mature walnut 7S globulin (indicating an N-terminus at amino acid 173 of the full sequence), the walnut 7S globulin preproprotein processing appears to be similar to that observed in cocoa and cotton, where a large hydrophilic N-terminal region is removed.^{11,24,26} Despite the well-documented presence of posttranslational processing of the 7S preproteins in various plants, the fates of the hydrophilic N-terminal regions are often unknown, as studies have focused primarily on characterization of the mature 7S proteins.

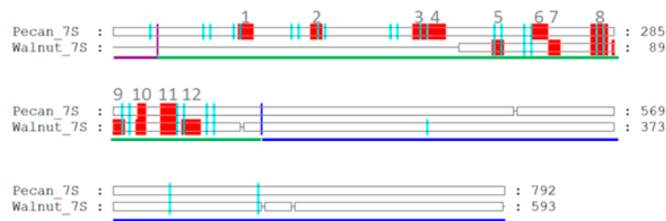
The peptide matches obtained from the LC-MS/MS analysis spanned the entire length of the N-terminal regions of the pecan and walnut 7S proprotein sequences, but the corresponding theoretical masses (46.3 and 21.8 kDa, respectively) are much larger than the molecular masses estimated for the 1D-PAGE bands (5.8–6.7 kDa), implying that this region has undergone significant proteolysis in the seed. MALDI-TOF MS was undertaken on the LMW pooled fraction (under nonreduced and reduced/alkylated conditions) to characterize the molecular masses of the 7S globulin N-terminal fragments and the 2S albumin (Figure 4). Due to the unfractionated nature of the LMW pool, the resulting spectra were complex. On the basis of the LC-MS/MS identifications from 1D-PAGE, the 2S albumins are responsible for peaks near m/z 12600 in the nonreduced sample and m/z 8400 and 3000–4000 in the reduced/alkylated sample. The 7S globulin N-terminal peptides were responsible for the remaining peaks in the m/z 5500–7500 region of both spectra.

Inspection of the m/z range corresponding to the 7S globulin N-terminal peptides (m/z 5500–7500) revealed a consistent pattern of peaks, which showed an m/z shift of +232 Da between the nonreduced and reduced/alkylated samples (Figure 4, inset; Table 3). A shift of this mass is equivalent to the reduction of two disulfide bonds and alkylation of four cysteine residues (net mass change of +58 Da per cysteine residue). The N-terminal regions of the walnut and pecan 7S proproteins are rich in cysteine residues (containing 12 and 25 residues, respectively), which appear to be arranged in clusters of 4 (Figure 3C).

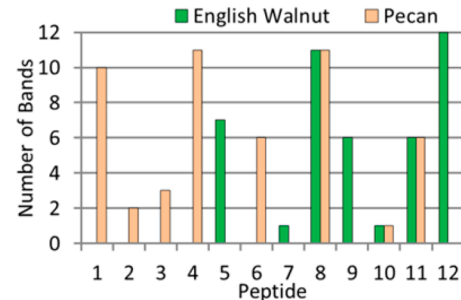
In addition to providing information about the 7S globulin N-terminal peptides, the MALDI-TOF spectra also revealed characteristics of the 2S albumin from walnut. The main peak corresponding to the intact (nonreduced) protein had a mass of 12617.2 Da (Figure 4). The predominant peaks for the small and large subunits were observed at $[M + H]^+$ 4640.1 and 8440.72 Da, respectively, which corresponds well with the predicted mass following reduction and alkylation of four cysteine bonds of the intact protein. Minor peaks were also observed in the regions associated with the intact 2S albumin (m/z 12373.25) and its subunits (m/z 8571.20, 8457.65, 4768.26, 4395.25, 4229.24, 4073.20, 4002.19, 3828.19, 3591.67, 3552.70, and 3435.28). Other tree nut 2S albumins, such as Brazil nut, have a documented level of heterogeneity, due to both sequence isoforms and posttranslational proteolytic processing.³² In the case of the 2S albumin from English walnut, however, it is difficult to assign all MALDI-TOF or LC-MS/MS spectra as there is only one sequence isoform present in the protein databases.

The role of the English walnut 2S albumin as an allergen was established previously,¹⁰ and it was given the allergen designation Jug r 1. The importance of the 7S N-terminal peptides as allergens in walnut is less well-understood. IgE immunoblotting of the LMW pooled fraction revealed that 3/

(A) Full 7S Globulin Sequences Diagram



(B) Frequency of Peptide Identification



(C) N-terminal Region of 7S Globulin Sequences

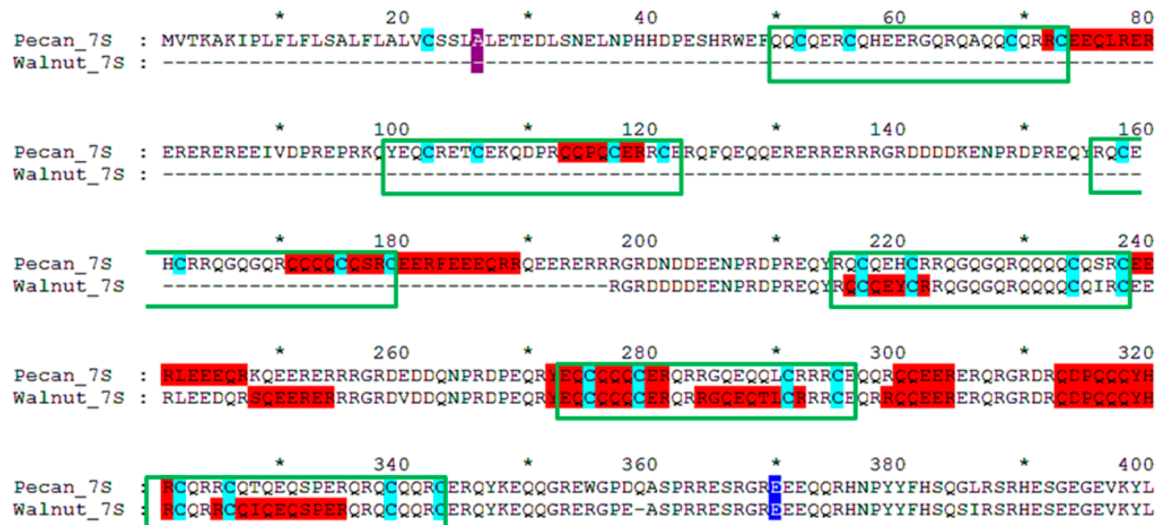


Figure 3. Peptide spectrum matches from anion exchange 1D-PAGE bands: (A) aligned, full-length English walnut and pecan 7S globulin sequences (Q9SEW4 and B3STU4, respectively) are represented as white bars, labeled with red compiled sequence coverage, cyan cysteine residues, purple predicted signal peptide, green N-terminal region, and blue mature 7S globulin protein.¹¹ (B) The frequency of peptide occurrence is indicated as the number of 1D-PAGE bands in which each peptide (as numbered in panel A) was identified. (C) N-terminal regions of aligned English walnut and 7S globulin sequences are shown with red compiled peptide spectrum matches from all anion exchange LC-MS/MS analysis, purple predicted end of signal peptide, blue N-terminus of mature peptide,¹¹ and green conserved four-cysteine motifs (C-X-X-X-C-X₁₀₋₁₂-C-X-X-X-C) labeled.

11 (27%) sera from walnut-sensitized individuals demonstrated IgE binding to the 7S N-terminal peptides, whereas 5/11 (45%) showed IgE binding to the 2S albumin (Figure 5).

The proportion of sera reacting to the 2S albumin was less than the 75% reported by Teuber et al., but the differences could be due to a number of factors, including differences in patient geographical distribution.¹⁰ Interestingly, in the current study, the IgE binding activity seems to be greatly decreased (to levels below the detection limit of the system utilized) when the proteins are analyzed under reducing conditions, which could indicate the presence of conformational epitopes. During the purification of nonspecific lipid transfer protein from English walnut, Pastorello et al. also isolated two peptides originating from the N-terminal region of the 7S vicilin-like globulin, which could be similar to the ones obtained in the present study based on N-terminal sequencing data.¹² In their study population, 10/46 (22%) patient sera showed IgE reactivity to at least one of the 7S globulin fragments, indicating similar levels of reactivity in the two populations.

The polypeptides from the N-terminal region of the English walnut 7S globulin proprotein described in this study are not an isolated occurrence. Similar cysteine-rich peptide fragments have been observed in macadamia nuts. Marcus et al. reported the presence of a series of antimicrobial peptides originating

from the N-terminal region of the macadamia nut 7S globulin proprotein.³³ All of the peptides contained the following four-cysteine motif: C-X-X-X-C-X₁₀₋₁₂-C-X-X-X-C. Both the walnut and pecan 7S globulin sequences contain the same motif repeated three and six times, respectively, throughout the N-terminal region (Figure 3C), and this type of four-cysteine motif is consistent with the mass shifts observed in the MALDI-TOF spectra. The cumulative evidence in the current study indicates that the walnut 7S N-terminal fractions contain a heterogeneous mixture of peptides centered around these types of four-cysteine motifs. The presence of these motifs provides valuable information about the structure and possible biological function of these peptides.

In addition to their documented presence in some tree nut species, similar peptides have been described in other plant species.³⁴⁻³⁶ The homologous peptide BWI-2b from buckwheat (*Fagopyrum esculentum*) functions as a trypsin inhibitor and exhibited IgE reactivity in 4/12 buckwheat-allergic subject sera.³⁴ In addition, the major buckwheat allergen, Fag e 3, also has homology with the 7S globulin proprotein N-terminal region.³⁵ Another homologous peptide from buckwheat, BWI-2c, was recently sequenced and described as having potent trypsin inhibition functionality and an α -helical hairpin structure, stabilized by disulfide bonds of the four-cysteine

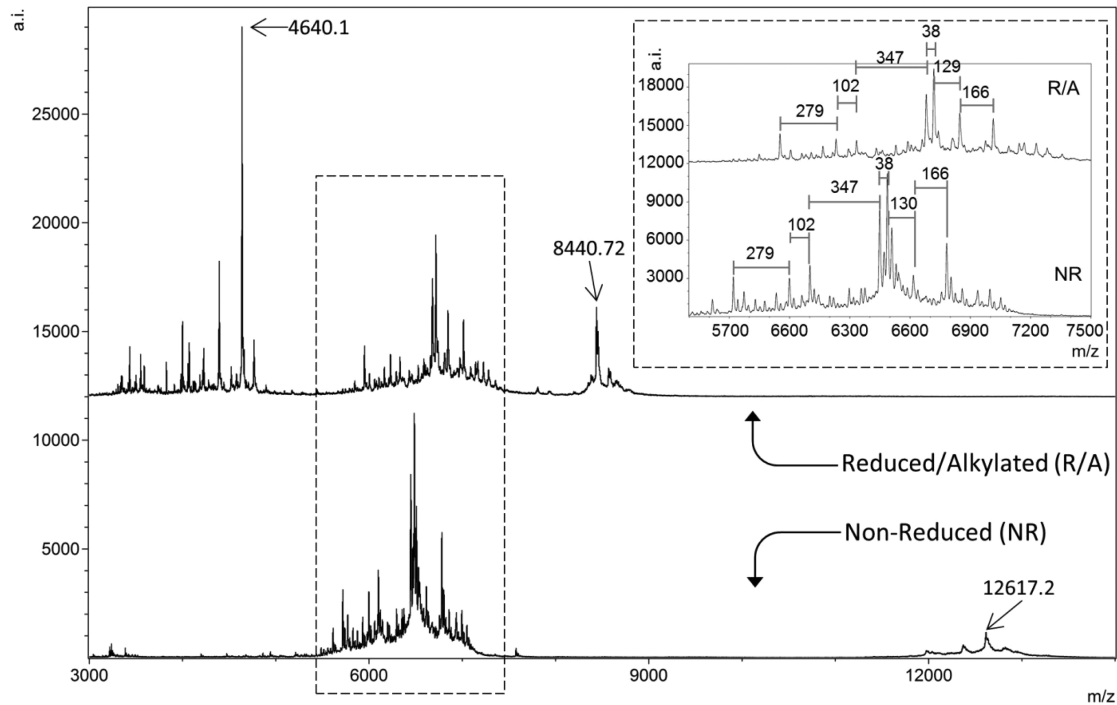


Figure 4. MALDI-TOF analysis of walnut LMW pool. Averaged spectra of the LMW gel filtration pool under nonreduced (NR) or reduced and alkylated (R/A) conditions are shown above, with average intensity (a.i.) displayed on the y-axis. Labels on inset spectra indicate conserved m/z differences between peaks.

Table 3. Mass Shifts of 7S Globulin N-Terminal Peptide MALDI-TOF Peaks Following Reduction and Alkylation^a

NR mean m/z	R/A mean m/z	$\Delta m/z$ (R/A-NR)
5720.2395	5952.9377	232.6981
5999.1986	6231.6017	232.4032
6101.4799	6333.9527	232.4728
6448.7647	6681.1299	232.3652
6486.8504	6718.6630	231.8126
6616.5919	6848.1240	231.5321
6782.3607	7014.4832	232.1224

^aValues shown indicate calculations of the m/z shift of the conserved pattern of peaks illustrated in Figure 4 (inset), under nonreduced (NR) and reduced and alkylated (R/A) conditions.

motif.³⁶ Finally, cotton (*Gossypium hirsutum*) also has a 7S globulin gene encoding for a hydrophilic N-terminal region (containing three repetitions of the four-cysteine motif), and an isolated heterogeneous population of peptides from this region demonstrated antifungal properties.³⁷ The independent description of similar N-terminal peptides derived from 7S globulin proprotein sequences in seeds from several different plant species implies that this type of peptide is produced consistently in a variety of plants. Due to the lack of complete plant genomes and proteomes, however, it is difficult to assess the full range of affected plant species.

Purification and analysis of the LMW, nonglycosylated fraction of walnut revealed the presence of two main polypeptide species: the 2S albumin and a mixture of

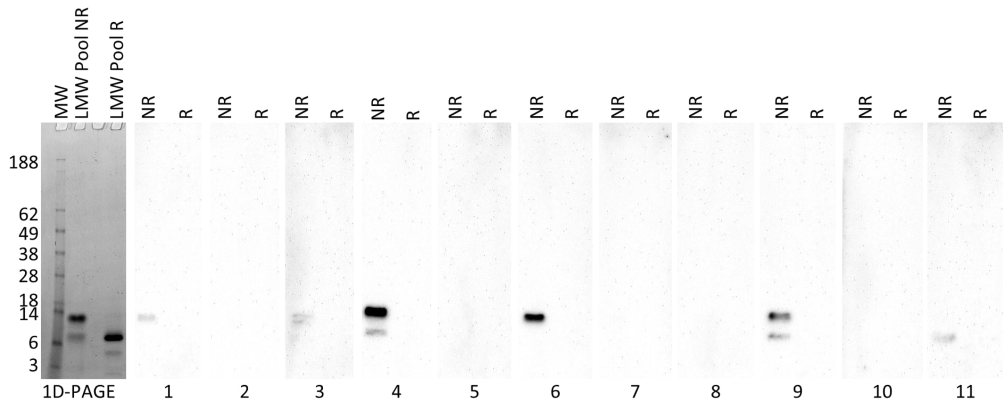


Figure 5. IgE reactivity to low molecular weight walnut proteins. Each IgE immunoblot includes the LMW pool (1.0 μg of protein) under nonreducing (NR) and reducing conditions (R), run in the same manner as on the 1D-PAGE gel replicate. Molecular weight markers (MW, SeeBlue prestained standards, 2 μL) were also run for each gel and blot. The numerical labels (1–11) below each immunoblot correspond to the individual serum samples from the walnut-allergic subjects described in Table 1.

polypeptides derived from the N-terminal region of the 7S seed storage globulin proprotein. N-terminal peptides derived from 7S globulin proprotein sequences and containing a conserved four-cysteine motif have been observed in a variety of plants, including walnuts, macadamia nuts, buckwheat, pumpkin, and cotton,^{12,24,27,33–37} and gene sequences suggest they may be produced in a number of other plants such as cocoa.²⁶ In cotton and pumpkin, there is quite convincing evidence that multiple protein products with separate functionalities are produced from the single proprotein sequence typically designated a 7S globulin.^{24,27,37} In cases when the peptides have been characterized, they seem to function primarily as a defense mechanism, with antimicrobial or trypsin inhibition activity.^{33,34,36,37}

Additionally, these four-cysteine motif containing peptides can be potential allergens, as illustrated by the IgE reactivity observed in walnut- or buckwheat-allergic subjects using immunoblot, ELISA, or RAST analysis.^{12,34,35} These data suggest that the N-terminal fragments could represent a subset of the 7S seed storage globulin allergens, which should be considered alongside other low molecular weight allergens such as 2S albumin and LTPs. However, as IgE reactivity alone is not fully indicative of biological activity, further in vivo or in vitro experiments (e.g., histamine release assays) should be conducted to confirm the clinical significance of the N-terminal peptides containing the four-cysteine motif. The implications for these peptides as potential allergens extend particularly to diagnostic and detection methods that aim to characterize allergens at the molecular level.

Component-resolved diagnosis using recombinant allergens, for example, could provide misleading or incomplete information if the N-terminal region of the 7S globulin proprotein is not analyzed separately from the mature protein, as these represent two independent allergenic components. In addition, the development of mass spectrometry-based methods for the detection of allergens in foods must properly assign peptide targets to particular proteins of interest and should recognize the independent nature of the 7S globulin proprotein N-terminal region. This study also shows that the 2S albumin from walnut exhibits a substantial amount of heterogeneity, which is not reflected in protein sequence databases. Without in-depth knowledge of protein isoforms and posttranslational processing, these types of heterogeneity make full characterization of native walnut proteins by mass spectrometry more challenging. As is the case with many plant foods, the genetic characterization of these species lags behind that of many animal and bacterial species. The continued expansion of genomic and proteomic data from plant foods is key to increasing our understanding of plant food allergens and will enhance our ability to use advanced techniques for characterization and quantitation of allergens in foods.

■ ASSOCIATED CONTENT

● Supporting Information

Table S1: Mascot Identification of Peptides and Proteins from LMW English Walnut Anion Exchange Bands. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*(M.L.D.) E-mail: mdowns2@unl.edu. Phone: (402) 472-5423. Fax: (402) 472-4474.

Funding

Research funding was provided by the Food Allergy Research & Resource Program at the University of Nebraska, a food industry-sponsored consortium of 80 food processing companies and their suppliers. This paper is independent research supported by the National Institute for Health Research Clinical Research Facility at University Hospital of South Manchester NHS Foundation Trust. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the Manchester Respiratory and Allergy Biobank (funded by the UK National Institute for Health Research) and the North West Lung Centre Charity for supporting this project. In addition we thank the study participants for their contribution. We also thank Francis Mulholland (Institute of Food Research, Norwich, UK) for mass spectrometry technical support and Dr. Claire Evers (Institute of Integrative Biology, University of Liverpool, UK) for helpful discussions.

■ ABBREVIATIONS USED

DWNE, defatted walnut flour; LMW, low molecular weight

■ REFERENCES

- (1) Osborne, T. B.; Campbell, G. F. Conglutin and vitellin. *J. Am. Chem. Soc.* **1896**, *18*, 609–623.
- (2) Osborne, T. B.; Harris, I. F. The globulin of the English walnut, the American black walnut and the butternut. *J. Am. Chem. Soc.* **1903**, *25*, 848–853.
- (3) Sicherer, S. H.; Muñoz-Furlong, A.; Godbold, J. H.; Sampson, H. A. US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. *J. Allergy Clin. Immunol.* **2010**, *125*, 1322–1326.
- (4) Ben-Shoshan, M.; Harrington, D. W.; Soller, L.; Fragapane, J.; Joseph, L.; St. Pierre, Y.; Godefroy, S. B.; Elliot, S. J. A population-based study on peanut, tree nut, fish, shellfish, and sesame allergy prevalence in Canada. *J. Allergy Clin. Immunol.* **2010**, *125*, 1327–1335.
- (5) Sicherer, S. H.; Furlong, T. J.; Muñoz-Furlong, A.; Burks, A. W.; Sampson, H. A. A voluntary registry for peanut and tree nut allergy: characteristics of the first 5149 registrants. *J. Allergy Clin. Immunol.* **2001**, *108*, 128–132.
- (6) Sampson, H. A.; Mendelson, L.; Rosen, J. P. Fatal and near-fatal anaphylactic reactions to food in children and adolescents. *N. Engl. J. Med.* **1992**, *327*, 380–384.
- (7) Pumphrey, R. S. H. Lessons for management of anaphylaxis from a study of fatal reactions. *Clin. Exp. Allergy* **2000**, *30*, 1144–1150.
- (8) Bock, S. A.; Muñoz-Furlong, A.; Sampson, H. A. Fatalities due to anaphylactic reactions to foods. *J. Allergy Clin. Immunol.* **2001**, *107*, 191–193.
- (9) Asero, R.; Antoniceilli, L.; Arena, A.; Bommarito, L.; Caruso, B.; Colombo, G.; Crivellaro, M.; De Carli, M.; Della Torre, E.; Della Torre, F.; Heffler, E.; Lodi Rizzini, F.; Longo, R.; Manzotti, G.; Marcotulli, M.; Melchiorre, A.; Minale, P.; Morandi, P.; Moreni, B.; Moschella, A.; Murzilli, F.; Nebiolo, F.; Poppa, M.; Randazzo, S.; Rossi, G.; Senna, G. E. Causes of food-induced anaphylaxis in Italian adults: a multi-centre study. *Int. Arch. Allergy Immunol.* **2009**, *150*, 271–277.
- (10) Teuber, S. S.; Dandekar, A. M.; Peterson, W. R.; Sellers, C. L. 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*, 807–814.

- (11) Teuber, S. S.; Jarvis, K. C.; Dandekar, A. M.; Peterson, W. R.; Ansari, A. A. 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. *J. Allergy Clin. Immunol.* **1999**, *104*, 1311–1320.
- (12) Pastorello, E. A.; Farioli, L.; Pravettoni, V.; Robino, A. M.; Scibilia, J.; Fortunato, D.; Conti, A.; Borgonovo, L.; Bengtsson, A.; Ortolani, C. Lipid transfer protein and vicilin are important walnut allergens in patients not allergic to pollen. *J. Allergy Clin. Immunol.* **2004**, *114*, 908–914.
- (13) Wallowitz, M.; Peterson, W. R.; Uratsu, S.; Comstock, S. S.; Dandekar, A. M.; Teuber, S. S. Jug r 4, a legumin group food allergen from walnut (*Juglans regia* Cv. Chandler). *J. Agric. Food Chem.* **2006**, *54*, 8369–8375.
- (14) Breiteneder, H.; Radauer, C. A classification of plant food allergens. *J. Allergy Clin. Immunol.* **2004**, *113*, 821–830.
- (15) Beyer, K.; Bardina, L.; Grishina, G.; Sampson, H. A. Identification of sesame seed allergens by 2-dimensional proteomics and Edman sequencing: seed storage proteins as common food allergens. *J. Allergy Clin. Immunol.* **2002**, *110*, 154–159.
- (16) Kleber-Janke, T.; Cramer, R.; Appenzeller, U.; Schlaak, M.; Becker, W.-M. Selective cloning of peanut allergens, including profilin and 2S albumins, by phage display technology. *Int. Arch. Allergy Immunol.* **1999**, *119*, 265–274.
- (17) Pastorello, E. A.; Farioli, L.; Pravettoni, V.; Spano, M.; Conti, A.; Ansaloni, R.; Rotondo, F.; Incorvaia, C.; Bengtsson, A.; Rivolta, F.; Trambaioli, C.; Previdi, M.; Ortolani, C. Sensitization to the major allergen of Brazil nut is correlated with the clinical expression of allergy. *J. Allergy Clin. Immunol.* **1998**, *102*, 1021–1027.
- (18) Pastorello, E. A.; Varin, E.; Farioli, L.; Pravettoni, V.; Ortolani, C.; Trambaioli, C.; Fortunato, D.; Giuffrida, M. G.; Rivolta, F.; Robino, A.; Calamari, A. M.; Lacava, L.; Conti, A. The major allergen of sesame seeds (*Sesamum indicum*) is a 2S albumin. *J. Chromatogr., B: Biomed. Sci. Appl.* **2001**, *756*, 85–93.
- (19) Rothham, J. M.; Wang, F.; Seamon, V.; Teuber, S. S.; Sathe, S. K.; Sampson, H. A.; Beyer, K.; Seavy, M.; Roux, K. H. Ana o 3, an important cashew nut (*Anacardium occidentale* L.) allergen of the 2S albumin family. *J. Allergy Clin. Immunol.* **2005**, *115*, 1284–1290.
- (20) Shewry, P. R.; Napier, J. A.; Tatham, A. S. Seed storage proteins: structures and biosynthesis. *Plant Cell* **1995**, *7*, 945–956.
- (21) Sordet, C.; Culerrier, R.; Granier, C.; Rance, F.; Didier, A.; Barre, A.; Rouge, P. Expression of Jug r 1, the 2S albumin allergen from walnut (*Juglans regia*), as a correctly folded and functional recombinant protein. *Peptides* **2009**, *30*, 1213–1221.
- (22) van Ree, R.; Cabanes-Macheteau, M.; Akkerdaas, J.; Milazzo, J.-P.; Loutelier-Bourhis, C.; Rayon, C.; Villalba, M.; Koppelman, S.; Aalberse, R.; Rodríguez, R.; Faye, L.; Lerouge, P. $\beta(1,2)$ -Xylose and $\alpha(1,3)$ -fucose residues have a strong contribution in IgE binding to plant glycoallergens. *J. Biol. Chem.* **2000**, *275*, 11451–11458.
- (23) Vizcaino, J. A.; Côté, R. G.; Csordas, A.; Dienes, J. A.; Fabregat, A.; Foster, J. M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; O'Kelly, G.; Schoenegger, A.; Ovelheiro, D.; Pérez-Riverol, Y.; Reisinger, F.; Ríos, D.; Wang, R.; Hermjakob, H. The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **2013**, *41*, D1063–D1069.
- (24) Chlan, C.; Pyle, J. B.; Legocki, A. B.; Dure, L., III. Developmental biochemistry of cottonseed embryogenesis and germination XVIII cDNA and amino acid sequences of members of the storage protein families. *Plant Mol. Biol.* **1986**, *7*, 475–489.
- (25) Doyle, J. J.; Schuler, M. A.; Godette, W. D.; Zenger, V.; Beachy, R. N.; Slightom, J. L. The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. Structural homologies of genes and proteins. *J. Biol. Chem.* **1986**, *261*, 9228–9238.
- (26) Spencer, M.; Hodge, R. Cloning and sequencing of a cDNA encoding the major storage proteins of *Theobroma cacao*. *Planta* **1992**, *186*, 567–576.
- (27) Yamada, K.; Shimada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I. Multiple functional proteins are produced by cleaving Asn-Gln Bonds of a single precursor by vacuolar processing enzyme. *J. Biol. Chem.* **1999**, *274*, 2563–2570.
- (28) Coates, J. B.; Medeiros, J. S.; Thanh, V. H.; Nielsen, N. C. Characterization of the subunits of β -conglycinin. *Arch. Biochem. Biophys.* **1985**, *243*, 184–194.
- (29) de Jong, E. C.; van Zijverden, M.; Spanhaak, S.; Koppelman, S. J.; Pellegroni, H.; Penninks, A. H. Identification and partial characterization of multiple major allergens in peanut proteins. *Clin. Exp. Allergy* **1998**, *28*, 743–751.
- (30) Wichers, H. J.; De Beijer, T.; Savelkoul, H. F. J.; van Amerongen, A. The major peanut allergen Ara h 1 and its cleaved-off N-terminal peptide; possible implications for peanut allergen detection. *J. Agric. Food Chem.* **2004**, *52*, 4903–4907.
- (31) Marsh, J.; Rigby, N.; Wellner, K.; Reese, G.; Knulst, A.; Akkerdaas, J.; van Ree, R.; Radauer, C.; Lovegrove, A.; Sancho, A.; Mills, C.; Vieths, S.; Hoffmann-Sommergruber, K.; Shewry, P. R. Purification and characterisation of a panel of peanut allergens suitable for use in allergy diagnosis. *Mol. Nutr. Food Res.* **2008**, *52*, S272–S285.
- (32) Moreno, F. J.; Jenkins, J. A.; Mellon, F. A.; Rigby, N. M.; Robertson, J. A.; Wellner, N.; Mills, E. N. C. Mass spectrometry and structural characterization of 2S albumin isoforms from Brazil nuts (*Bertholletia excelsa*). *Biochim. Biophys. Acta* **2004**, *1698*, 175–186.
- (33) Marcus, J. P.; Green, J. L.; Goulter, K. C.; Manners, J. M. A family of antimicrobial peptides is produced by processing of a 7S globulin protein in *Macadamia integrifolia* kernels. *Plant J.* **1999**, *19*, 699–710.
- (34) Park, S.-S.; Abe, K.; Kimura, M.; Urisu, A.; Yamasaki, N. Primary structure and allergenic activity of trypsin inhibitors from the seeds of buckwheat (*Fagopyrum esculentum* Moench). *FEBS Lett.* **1997**, *400*, 103–107.
- (35) Choi, S.-Y.; Sohn, J.-H.; Lee, Y.-W.; Lee, E.-K.; Hong, C.-S.; Park, J.-W. Characterization of buckwheat 19-kD allergen and its application for diagnosing clinical reactivity. *Int. Arch. Allergy Immunol.* **2007**, *144*, 267–274.
- (36) Oparin, P. B.; Mineev, K. S.; Dunaevsky, Y. E.; Arseniev, A. S.; Belozersky, M. A.; Grishin, E. V.; Egorov, T. A.; Vassilevski, A. A. Buckwheat trypsin inhibitor with helical hairpin structure belongs to a new family of plant defence peptides. *Biochem. J.* **2012**, *446*, 69–77.
- (37) Chung, R. P. T.; Neumann, G. M.; Polya, G. M. Purification and characterization of basic proteins with in vitro antifungal activity from seeds of cotton, *Gossypium hirsutum*. *Plant Sci.* **1997**, *127*, 1–16.
- (38) Simons, F. E. R.; Arduoso, L. R. F.; Bilò, M. B.; El-Gamal, Y. M.; Ledford, D. K.; Ring, J.; Sánchez-Borges, M.; Senna, G.; Sheikh, A.; Thong, B. Y. World Allergy Organization anaphylaxis guidelines: summary. *J. Allergy Clin. Immunol.* **2011**, *127*, 587–593.

SUPPORTING INFORMATION

Table S1: Mascot Identification of Peptides and Proteins from LMW English Walnut Anion Exchange Bands.

All peptides assigned to English walnut (*J. regia*; EWN) or pecan (*C. illinoensis*; PEC) proteins are given in the table. All of the pecan 7S vicilin protein identifications designated as accession number B3STU4 also gave identical results for another pecan 7S sequence isoform, B3STU7. Individual peptides with an expectation value less than 0.05 are statistically significant peptide spectrum matches.

Band	Rank	Protein	UniProt Accession	Protein Score	Sequence Coverage	Unique Peptides	Peptide	Expectation Value
AE1	1	EWN 2S	P93198	366	40%	7	DLPNECGISSQR	0.000027
							GEEMEEMVQSAR	0.00000013
							QQNLNHCQYYLR	0.023
							QQQQQGLR	0.023
							QQQQQGLRGEEMEEMVQSAR	0.000095
							RQQQQQGLR	0.0062
							SGGYDEDNQR	0.000059
	11	EWN LTP	C5H617	35	7%	1	CGVSVPYK	0.03
AE2	1	PEC 7S	B3STU4	239	7%	10	CEEQLR	0.2
							CEEQLRER	5.5
							CEERFEEEQR	0.011
							CEERLEEEQR	0.1
							FEEEQR	0.045
							FEEEQRR	1.1
							QDPQQQYHR	11
							QQQQCQSR	1.5
							RCEEQLR	0.19
							YEQCQQQCER	0.00013
	2	EWN 7S	Q9SEW4	175	7%	5	CQIQEQSPER	0.0039
							GQEQTLCR	0.0016
							QCQEYCR	0.0036
							QDPQQQYHR	11
							YEQCQQQCER	0.00013
	4	EWN 2S	P93198	59	23%	2	DLPNECGISSQR	0.0024
							QQQQQGLRGEEMEEMVQSAR	1
AE3	1	PEC 7S	B3STU4	151	4%	6	CEEQLR	0.077
							CEEQLRER	2.1
							CEERFEEEQR	0.025
							FEEEQR	4.1
							RCEEQLR	0.006
							YEQCQQQCER	0.014
	3	EWN 7S	Q9SEW4	66	5%	3	QCQEYCR	0.0064
							RCQIQEQSPER	4.1
							YEQCQQQCER	0.014
AE4	1	PEC 7S	B3STU4	200	4%	8	CEEQLR	0.19
							CEEQLRER	7.9
							CEERFEEEQR	11
							FEEEQR	0.31
							LEEEQR	0.063
							QQQQCQSR	0.038
							RCEEQLR	0.0054
							YEQCQQQCER	0.0035
	2	EWN 7S	Q9SEW4	114	5%	3	CQIQEQSPER	0.0056
							QCQEYCR	0.00033
							YEQCQQQCER	0.0035
AE5	1	EWN 7S	Q9SEW4	224	8%	7	CQIQEQSPER	0.00068
							GQEQTLCR	0.0056
							QCQEYCR	0.0097
							QDPQQQYHR	19
							RCQIQEQSPER	0.012

[illegible]

[illegible]

[illegible]