

Size-Selective Fractionation and Visual Mapping of Allergen Protein Chemistry in *Arachis hypogaea*

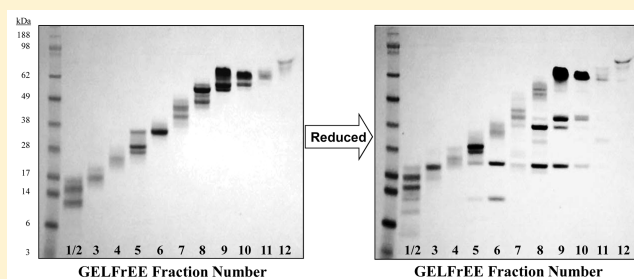
Christine M. Hebling,* Mark M. Ross, John H. Callahan, and Melinda A. McFarland

U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition, 5100 Paint Branch Parkway, College Park, Maryland 20740, United States

S Supporting Information

ABSTRACT: Peanuts (*Arachis hypogaea*) in addition to milk, eggs, fish, crustaceans, wheat, tree nuts, and soybean are commonly referred to as the “big eight” foods that contribute to the majority of food allergies worldwide. Despite the severity of allergic reactions and growing prevalence in children and adults, there is no cure for peanut allergy, leaving avoidance as the primary mode of treatment. To improve analytical methods for peanut allergen detection, researchers must overcome obstacles involved in handling complex food matrices while attempting to decipher the chemistry that underlies allergen protein interactions. To address such challenges, we conducted a global proteome characterization of raw peanuts using a sophisticated GELFrEE-PAGE-LC-MS/MS platform consisting of gel-based protein fractionation followed by mass spectrometric identification. The in-solution mass-selective protein fractionation: (1) enhances the number of unique peptide identifications, (2) provides a visual map of protein isoforms, and (3) aids in the identification of disulfide-linked protein complexes. GELFrEE-PAGE-LC-MS/MS not only overcomes many of the challenges involved in the study of plant proteomics, but enriches the understanding of peanut protein chemistry, which is typically unattainable in a traditional bottom-up proteomic analysis. A global understanding of protein chemistry in *Arachis hypogaea* ultimately will aid the development of improved methods for allergen detection in food.

KEYWORDS: peanut allergens, Ara h, GELFrEE, disulfide bonds, mass spectrometry



■ INTRODUCTION

Peanut allergy is a human sera immunoglobulin E (IgE)-mediated reaction that affects greater than 1% of children in the United States.¹ Although several approaches to effectively treat food allergy are under study, the current therapy for food allergy is limited to a strict avoidance of the offending food. This approach is generally effective, but compliance can be difficult due to cross contamination during processing and food labeling errors. To aid in the development of improved allergen detection methods, mass spectrometry has become an emerging field in food science and technology. This field, termed allergenomics,^{2,3} has proven to be essential at various levels in the study of food allergens, from structural characterization to quantitative detection in a variety of food matrices.

To date, the World Health Organization (WHO) and International Union of Immunological Societies (IUIS) Allergen Nomenclature Subcommittee have identified 10 peanut proteins as official allergens, capable of eliciting allergic reactions and/or being recognized by IgE antibodies from peanut allergic patients (www.allergen.org). Representing the majority of the total protein in peanuts, globulins (Ara h 1 and Ara h 3/4) and 2S albumins (Ara h 2, Ara h 6, and Ara h 7) are the allergens responsible for eliciting immune responses in the majority of peanut-allergic individuals.^{4–6} Protein globulins Ara h 1 (7S) and Ara h 3/4 (11S) are seed storage proteins that

belong to the cupin superfamily. The Ara h 1 protein allergen is characterized as a ~70 kDa glycoprotein that occurs naturally as a noncovalent trimer capable of aggregating into larger assemblies.^{4,7} Ara h 3/4, on the other hand, associates into a heat-stable hexameric structure and is characterized by multiple homologous protein isoforms.^{8–11} In accordance with other legume storage proteins, Ara h 3/4 is proteolytically cleaved at a conserved asparagine–glutamine bond, yielding disulfide-linked acidic and basic subunits.^{12–14} Continued post-translational processing of Ara h 3/4 results in a complex series of low molecular weight polypeptide fragments.^{9,15,16} While a functional explanation behind the presence of multiple protein isoforms and extensive protein processing for legume 11S globulins is not yet known, the development of analytical methods to investigate plant protein chemistries is warranted for improved characterization of Ara h 3/4 allergenicity.

Likely the most potent peanut allergen, Ara h 2 is recognized by the IgE of >90% of peanut allergic patients.^{5,6,17} Ara h 2 and two minor peanut allergens (Ara h 6 and Ara h 7) are structurally distinguished by conserved cysteine-containing subsequences and associated disulfide bonds.¹⁸ Ara h 2 is characterized as having two classes of protein isoforms (2.01

Received: July 9, 2012

Published: October 1, 2012

and 2.02) that vary by the addition of a 12-residue sequence in the flexible loop region of the protein structure.¹⁹ Disulfide bond assignments, post-translational modifications, and sequence variants between the four identified isoforms have been characterized for Ara h 2.²⁰ In addition to the major peanut allergens listed above, actin-binding profilin (Ara h 5),²¹ Bet v 1-related pathogenesis-related protein (Ara h 8),²² nonspecific lipid transfer protein (Ara h 9),^{23,24} and oil body oleosins (Ara h 10 and Ara h 11)²⁵ have been described and evaluated as minor allergens in *Arachis hypogaea*.

Compared to model organisms such as yeast (*Saccharomyces cerevisiae*) and humans, proteomic research in plants has not advanced at the same rate. The lack of availability of a nonredundant and accurately annotated *A. hypogaea* genome further complicates the comprehensive identification of proteins and corresponding isoforms. In addition to challenges in data interpretation, high-throughput sample preparation methods are limited by complexities inherent in food matrices. Plant materials characteristically include cell wall and storage polysaccharides, lipids, and phenolic compounds that interfere with protein extraction and sample cleanup. Furthermore, the most prominent peanut allergens, classified as seed storage proteins, exhibit a propensity to aggregate, a tendency for post-translational modification, and a resistance to proteolytic digestion, which limit the use of solution-based fractionation methodologies (e.g., size-exclusion chromatography)⁹ for the investigation of intact protein chemistries. In an effort to better understand the structural and functional interactions among *A. hypogaea* protein allergens, a mass-selective gel-assisted protein fractionation platform was designed through the combination of electrophoretic and chromatographic techniques with peptide identification by mass spectrometry.

Previous work utilizing two-dimensional gel-electrophoresis and mass spectrometry paved the road for proteomics investigations of *A. hypogaea* allergens in varying genotypes and cultivars.^{26–28} While informative, two-dimensional gels are visually complex and analytically cumbersome in the identification of individual protein bands. Combining the advantages of gel-based protein fractionation with the need for solution-based separations, a commercial technique founded on the concept of tube-gel electrophoresis^{29,30} was developed in 2008 by Tran and Doucette³¹ to allow for electrophoretic size-based protein fractionation. This technique, termed gel-eluted liquid fraction entrapment electrophoresis, or GELFrEE, is modeled after the traditional one-dimensional preparative SDS-PAGE separation where protein samples are solubilized in SDS and separated by molecular weight into liquid fractions depending on the duration and electrophoretic field applied to the separation.^{31,32} For practical applications and proteomic challenges, GELFrEE has provided a fractionation platform for molecular weight-based profiling of protein extracts coupled to top-down^{33–35} and bottom-up proteomics.^{32,36}

In the current study, a GELFrEE-polyacrylamide gel electrophoresis (PAGE)-liquid chromatography (LC)-tandem mass spectrometry (MS/MS) platform (GELFrEE-PAGE-LC-MS/MS) was developed to provide a global map of the *A. hypogaea* proteome. Using a combination of analytical techniques, GELFrEE-PAGE-LC-MS/MS introduces a multi-dimensional platform for global protein characterization with several advantages compared with those of traditional bottom-up proteomics. In-solution fractionation by GELFrEE allows for mass-based intact protein separation that, when combined with trypsin digestion and LC-MS/MS, yields improved

protein coverage and differentiation of protein isoforms. Isolation of disulfide-linked protein complexes, furthermore, can be achieved visually by SDS-PAGE and Western blot analyses of unreduced and reduced GELFrEE protein fractions. Aligning protein fractions with nominal protein molecular weight ranges offers valuable information regarding protein interactions and structural rearrangements that can be interrogated by LC-MS/MS for molecular identification. The elucidation of peanut protein chemistry promotes improved detection of peanuts in complex food matrices while providing new insights into peanut protein allergenicity.

■ EXPERIMENTAL SECTION

Materials

All chemicals were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Raw Runner peanuts were provided by the Golden Peanut Company (Alpharetta, GA). Sample cleanup was facilitated with a proprietary BioRad ReadyPrep precipitation kit (Hercules, CA). GELFrEE 8100 cartridges and supplies were obtained from Expedeon Protein Discovery (San Diego, CA). NuPAGE gels, gel reagents, and iBlot Western blotting supplies were purchased from Invitrogen Life Technologies (Grand Island, NY). Polyclonal chicken anti-Ara h 1, chicken anti-Ara h 2, and chicken anti-Ara h 3 antibodies (IgY, custom synthesized by Sigma Immunosys, The Woodlands, TX)^{26,37} in addition to purified Ara h 1 and Ara h 2 protein standards isolated from raw peanut extracts^{38,39} were graciously provided by Dr. Soheila Maleki (U.S. Department of Agriculture). Thermo Scientific Pierce (Rockford, IL) Super-Signal West Pico enhanced chemiluminescent (ECL) substrate was purchased for detection of proteins in Western blot applications. Estimated protein concentrations were determined using a Qubit Fluorometer (Invitrogen Life Technologies, Grand Island, NY). All sample digestions were performed with Porcine Trypsin (Promega, Madison, WI) with or without the presence of Waters acid-labile surfactant RapiGest SF (Milford, MA). Optima grade solvents for liquid chromatographic analysis were purchased from Fisher Scientific (Fair Lawn, NJ).

Protein Isolation from Raw Peanuts

Peanut kernels were frozen with liquid nitrogen, skinned, and ground to a coarse flour in a stainless steel blender. To obtain a fine powder, the peanut flour was frozen with liquid nitrogen and milled using an electronic mortar and pestle. The peanut flour was defatted using a 1:10 (w/v) flour/hexane mixture in a glass conical vial for 30 min while rotating. The hexane layer was decanted, and the defatting protocol was repeated two additional times. The defatted peanut flour was vacuum dried.

Defatted peanut flour (50 mg) was extracted with 1000 μ L of 10 mM phosphate buffered saline (PBS) pH 7.4, 25 mM iodoacetamide, and a plant protease inhibitor cocktail. The protease inhibitor solution was prepared in dimethyl sulfoxide (DMSO) and contained the following components: 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), bestatin, pepstatin A, E-64, leupeptin, and 1,10-phenanthroline. The solubilized protein mixture was vortexed for 15 min, sonicated in a water bath for 15 min at 4 °C, and allowed to rotate end-over-end for 3 h at room temperature. Centrifugation at 12,000 $\times$ g pelleted the undissolved peanut debris. The protein content of the extraction supernatant was measured using a Qubit Fluorometer. Protein extracts were stored temporarily at 4 °C.

GELFrEE Separations

Details on the operation of the GELFrEE system can be obtained from Tran et al.^{31,32} In brief, 300 μ g of peanut protein extract (described above) were precipitated using a BioRad ReadyPrep kit and reconstituted in Tris-acetate GELFrEE sample buffer. The sample was inserted into the loading chamber of a 10% Tris-acetate GELFrEE cartridge, while HEPES electrode running buffer (pH 7.8) was added to the anode, cathode, and sample collection chambers. At user-defined intervals, the content in the sample collection chamber was manually transferred to a new sample vial. To ensure proper buffering capacity, the running buffer was replaced in the anode and cathode buffer reservoirs after approximately 30 min of cumulative run time at 100 V or 60 min of cumulative run time at 50 V. In total, samples were partitioned into 12 fractions over 160 min at room temperature. A duplicate sample extract was analyzed in parallel as a biological replicate to determine fractionation and mass spectrometric identification reproducibility.

SDS-PAGE and Western Blot Analysis

GELFrEE fractions were analyzed by discontinuous SDS-PAGE on NuPAGE Novex gradient 4–12% Bis-Tris gels with NuPAGE MES [2-(*N*-morpholino)ethanesulfonic acid] SDS running buffer. GELFrEE fractions were prepared under unreduced or reduced (addition of 50 mM dithiothreitol) sample conditions prior to gel analysis. PAGE gels were stained using SimplyBlue SafeStain (Coomassie) and imaged on a Kodak Image Station 2000R. For protein molecular weight referencing, the SeeBlue Plus2 prestained protein ladder was loaded on each gel.

The iBlot Western Blotting device was utilized to transfer proteins from the SDS-PAGE gel to a PVDF membrane (0.2 μ m). After transfer, membranes were blocked for 1 h at room temperature with a 5% nonfat dry milk solution in 5 mM Tris-buffered saline with 0.05% Tween 20 (TBST). Chicken anti-Ara h 1 (1:5000), chicken anti-Ara h 2 (1:8000), and chicken anti-Ara h 3 (1:5000) antibodies were diluted (v/v) in TBST containing 5% nonfat dry milk. The blocked membrane was incubated with each primary antibody (individually) for 1 h at room temperature. The membrane was washed (3 $\times$ 5 min) with TBST and incubated for one hour at room temperature with the secondary rabbit anti-chicken/turkey HRP antibody (1:3000). Following secondary antibody incubation, membranes were washed with TBST and incubated with an enhanced chemiluminescence (ECL) substrate according to the manufacturer's instructions. The chemiluminescence was measured using a Kodak Image Station 2000R.

Sample Cleanup and Digestion: GELFrEE

Detergent removal from individual GELFrEE fractions was facilitated using Pierce detergent removal spin columns (0.5 mL resin volume). Prior to in-solution digestion, sample fractions were concentrated using pretreated Millipore Amicon molecular weight cutoff filters (YM-3). The concentrated fractions were eluted in 50 mM ammonium bicarbonate and the protein content of each GELFrEE fraction was measured using a Qubit Fluorometer.

In-solution tryptic digestion of GELFrEE fractions were accomplished after sample reduction (10 mM dithiothreitol) and alkylation (25 mM iodoacetamide) utilizing a 1:25 ratio of proteolytic enzyme to total protein. Digestion solutions were prepared in 50 mM ammonium bicarbonate for overnight (12–14 h) digestion at 37 °C and were quenched to a final

concentration of 2% acetonitrile and 0.2% formic acid in a total volume of 50 μ L.

For in-gel digestions, gel bands were excised and placed in an Eppendorf tube pretreated with 50% acetonitrile in Optima grade water. As a control, a gel band free of loaded protein was also processed to provide a background sample for digested gel band analyses. The gel bands were destained with 40% ethanol in 50 mM ammonium bicarbonate, washed with 50 mM ammonium bicarbonate, and cleaved into <1 mm³ pieces prior to dehydration with 100% acetonitrile. To the dried protein gel piece was added 30 μ L of trypsin proteolytic digestion solution (100 ng/ μ L) to reswell the gel band on ice for 45 min. An overlay of 10% acetonitrile in 40 mM ammonium bicarbonate was added, and digestion was carried out at 37 °C overnight. After digestion, the supernatant was removed and the remaining gel band was extracted with 60% acetonitrile in 40 mM ammonium bicarbonate for 30 min while vortexing. The extraction supernatant was removed and lyophilized to dryness before combining with the digestion supernatant for liquid chromatographic analysis.

Sample Cleanup and Digestion: Whole Lysate

Raw peanut protein extract (30 μ g) was precipitated using a BioRad ReadyPrep kit. The protein pellet was reconstituted in RapiGest SF acid-labile surfactant at a final concentration of 0.1% in 50 mM ammonium bicarbonate and heated for 5 min. The samples were reduced and alkylated with dithiothreitol and iodoacetamide at 10 and 25 mM final concentrations, respectively. Samples were digested at a 1:25 trypsin to total protein concentration overnight at 37 °C. The digestions were quenched with the addition of 0.5% TFA and incubated for 1 h at 60 °C to degrade the surfactant. Samples were centrifuged and passed through a Pierce C18 spin column for cleanup. Internal standards, Yeast Alcohol Dehydrogenase (10 pmol), Yeast Enolase (2 pmol), and Rabbit Phosphorylase B (25 fmol), were added to each sample as process, digest, and injection quality control internal standards, respectively.

Liquid Chromatography and Tandem Mass Spectrometry

An AB Sciex Eksigent 2D-nanoLC Ultra (Dublin, CA) was operated under 1-D dual-pump trapping mode where samples were loaded using a metered 2500 nL injection onto a Thermo Dionex Acclaim PepMap 100 trapping column (0.5 cm $\times$ 300 μ m i.d.; 5 μ m C18 particles, 100 Å) at 5 μ L/min with 97% Optima grade water, 3% acetonitrile, and 0.1% formic acid. Reverse phase separation was completed on a New Objective PicoFrit analytical capillary column (10 cm $\times$ 75 μ m i.d.) packed with 2.7 μ m C18 Halo particles (90 Å) and a 15 μ m tip. Column flow rate was maintained at 400 nL/min. Peptides from GELFrEE fractions were eluted with a linear gradient of 3–40% B in 30 min (50 min total run time) where mobile phase A was prepared with 0.1% (v/v) LC–MS grade formic acid in Optima grade water and mobile phase B with 0.1% (v/v) LC–MS grade formic acid in Optima grade acetonitrile. For lysate samples, the gradient was a slightly modified step gradient of 3–30% B in 65 min and 30–40% B from 65 to 80 min (100 min total run time).

Mass spectrometry analysis was accomplished using a Thermo LTQ-Orbitrap XL instrument with dynamic exclusion data dependent acquisition enabled. A high-resolution full FTMS profile spectrum (scan range 400–2000 *m/z*; resolving power 60,000; 1 microscan) was acquired in parallel with MS/MS centroid spectra for eight of the most abundant precursor ions. Collision-induced dissociation (CID) mass spectra were

acquired in the LTQ at one microscan per spectrum with a signal threshold of 1×10^3 counts and normalized collision energy of 35% (isolation width 2.0 m/z). Automatic gain control allowed accumulation of up to 5×10^5 ions for FTMS full scans and 1×10^4 ions for ion trap MS/MS (ITMSn) scans with a maximum injection time of 100 ms. Dynamic exclusion of previously selected ions was applied for 15 s with a repeat count of 2, a repeat duration of 15 s, and exclusion mass width of 0.6 Da with unassigned and 1+ charge states rejected. Biological sample replicates were prepared and analyzed in technical replicates (GELFrEE $\times 2$; whole lysate $\times 3$).

Database searches of MS/MS spectra were performed using a Mascot v 2.3 (Matrix Science, London, U.K.) search engine against a custom *A. hypogaea* database merged with a Swiss-Prot human protein sequence library and common protein contaminants (24,284 sequences). The database search parameters were specified as full trypsin enzyme digestion with up to two missed cleavages, fixed cysteine carbamidomethylation, variable hydroxyproline, oxidation of methionine, and deamidation of asparagine and glutamine. Peptide and protein level parsimony comparisons across multiple experiments were evaluated using MassSieve v. 1.12.⁴⁰ For peptide and protein identification, only assigned peptide sequences with Mascot ion scores greater than or equal to the identity score threshold (at $p < 0.05$) were classified as confident matches. Peptides required identification in both technical and biological replicates with a minimum of two confident peptide matches per protein. Using data analysis filters and criteria specified above, peptide identifications correspond to a 5% false discovery rate (FDR). Indeterminate identifications, or peptides whose MS/MS spectra were matched to a secondary sequence with an equivalent score, were retained throughout the sample analysis or grouped with the protein sequence providing the greatest support for identification. For experimental method comparisons, unique peptides were collapsed on the basis of charge state and modification.

Spectral counts, or the total number of MS/MS spectra assigned to a given peptide with ion scores greater than identity, were utilized to provide a measure of relative peptide abundance between data sets. To correct for instrumental inconsistencies between replicate injections, peptide spectral counts were normalized on the basis of the total number of MS/MS scans. For most sample analyses, peptides represented by a single spectral count were removed. More stringent conditions were used to simplify the assignment of Ara h 3/4 isoforms (Figure 5), reporting only peptides with greater than two MS/MS scans (spectral counts). For all studies, protein contaminants and internal standards were not included in the data analysis.

Due to the high degree of sequence homology between *A. hypogaea* protein isoforms, peptides were clustered into protein families to minimize redundancy. A detailed report of assigned protein clusters with associated protein family names is outlined in Supplemental Table 1. In addition to protein clustering, parsimony type (discrete, differentiable, superset, or equivalent),⁴⁰ the number of confidently assigned peptides, and percent coverage are reported for each protein. Subset and subsumable proteins were excluded from analysis in the body of the manuscript.

RESULTS AND DISCUSSION

Arachis hypogaea Protein Mapping by Size-Selective GELFrEE Fractionation

Peanut allergens have been extensively characterized over the past decade, covering a range of biochemical experimentation including, but not limited to, immunochemical response, molecular characterization, and structural determination.⁴¹ Mass spectrometry-driven initiatives have recently emerged in food science providing sequence-specific information for structural investigation of protein allergens and peptide quantification. A traditional means for comparative protein determination is an analysis of the unfractionated whole peanut lysate. A one-dimensional SDS-PAGE gel (Figure 1A) depicts

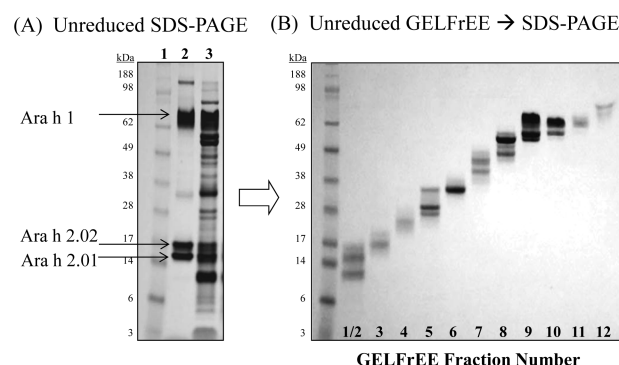


Figure 1. Evaluating size-based separations of PBS-extracted raw peanut by SDS-PAGE. (A) One-dimensional SDS-PAGE gel of unreduced whole peanut extract. Lane 1: SeeBlue Plus2 prestained molecular weight ladder; Lane 2: purified peanut allergen protein standards Ara h 1 (~70 kDa) and Ara h 2.01/2.02 (16–18 kDa); Lane 3: PBS extract of raw peanut. (B) One-dimensional SDS-PAGE gel of unreduced 10% Tris-acetate GELFrEE fractions for PBS-extracted raw peanut. GELFrEE fractions were collected in sequential increasing molecular weight intervals. A SeeBlue Plus2 prestained protein ladder is provided for molecular weight reference.

the electrophoretic protein banding patterns for purified Ara h 1 (lane 2), purified Ara h 2 (lane 2), and a complete PBS extraction of raw peanut (lane 3) under unreduced sample conditions. While protein standards (lane 2) can be used to discern specific protein bands in the whole peanut extract (lane 3), the complexity of an unfractionated peanut lysate makes it difficult to discriminate and identify individual protein components.

To visually simplify the banding pattern of a one-dimensional whole lysate electrophoretic separation, a gel-aided fractionation methodology (GELFrEE) was utilized to provide mass-selective separation of proteins in a raw peanut extract over the molecular weight range of 3.5–100 kDa. Individual protein fractions isolated from a GELFrEE separation of unreduced raw peanut flour extract can be visualized on a one-dimensional SDS-PAGE gel as depicted in Figure 1B. As compared to unfractionated whole protein techniques, the GELFrEE platform provides a tunable method by which electrophoretic potentials and fraction collection durations can be optimized for a given sample. Complementing intact protein molecular weights obtained by SDS-PAGE, individual GELFrEE fractions can be investigated by in-solution LC–MS/MS.

Unreduced raw peanut GELFrEE eluted protein fractions were digested by trypsin and analyzed using a bottom-up

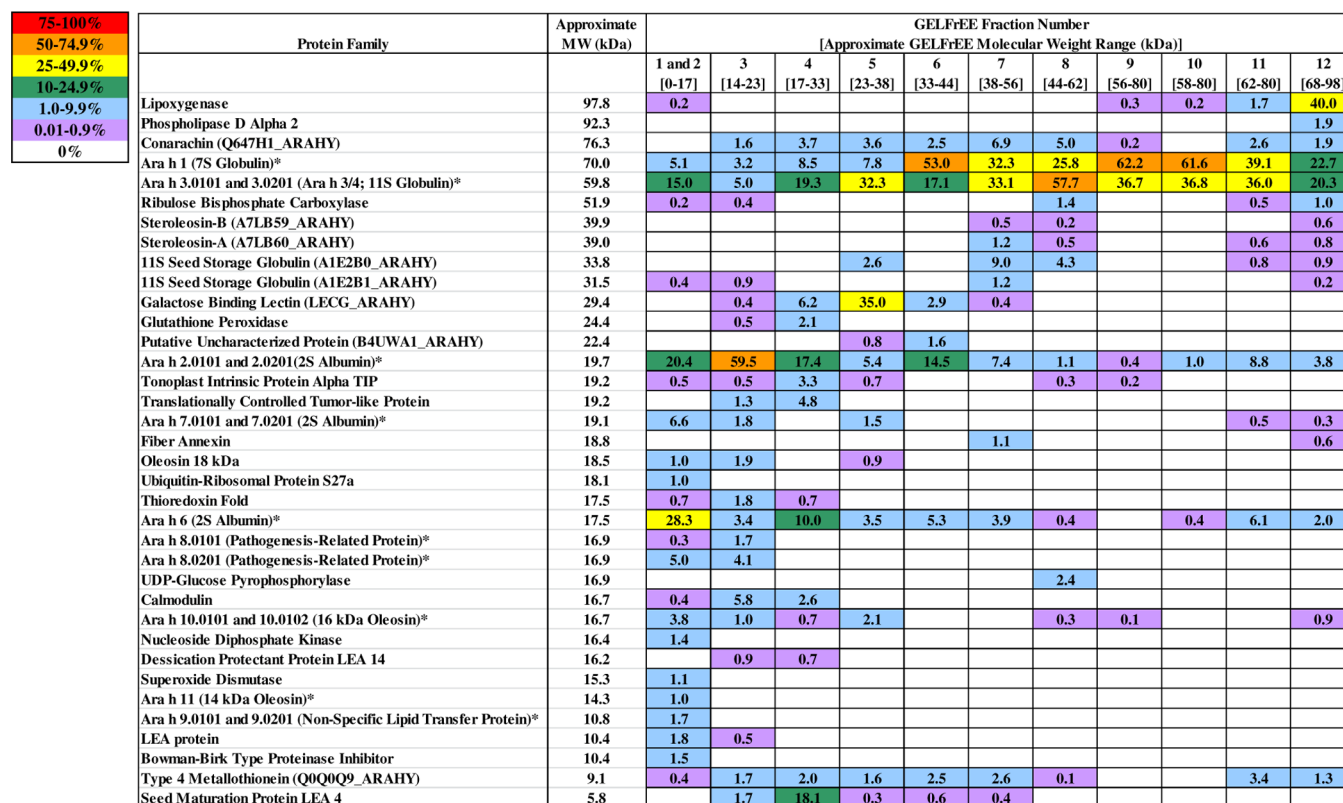


Figure 2. GELFrEE-LC-MS/MS platform for proteome visualization. Visual mapping of protein family relative abundance as determined by peptide spectral counts in trypsin-digested GELFrEE fractions. Protein families are sorted according to average intact molecular weight as reported in the customized SwissProt/TrEMBL database. Normalized abundances for each protein family (based on MS/MS spectral counts) are represented as a percentage of the total number of MS/MS scans for each GELFrEE fraction. Percentage values are classified into categories based upon the color scale provided. GELFrEE fractions are associated with approximate electrophoretic molecular weight ranges to estimate protein migration. *Arachis hypogaea* protein allergens are indicated by an asterisk. All known peanut allergen classes were identified; however, Ara h 5, or profilin, was identified at a percentage level below the 1% threshold to be displayed in this table.

proteomics approach. GELFrEE-LC-MS/MS of individual protein fractions generated a visual proteome map enabling protein tracking within a biological system based upon molecular weight (Figure 2). To mimic the format of a SDS-PAGE gel, proteins are aligned in descending order of molecular weight with labeled GELFrEE fractions along the horizontal axis. Relative protein abundance, as determined by spectral counts, is represented as a percentage of the total spectral counts for each GELFrEE fraction. Percentage values are displayed for each protein identified within a GELFrEE fraction and associated with a colorimetric intensity in descending order from red to violet with white indicating absence of protein identification. Within a given GELFrEE fraction (vertical column), a hierarchical ranking of protein abundance allows comparisons between predicted protein molecular weight and electrophoretic gel migration. Since the total protein content is not equal between GELFrEE fractions, fraction-to-fraction comparisons (horizontal rows) are limited to qualitative relative protein abundances. Colorimetric relationships between GELFrEE fractions exhibit a bias based upon the total protein loaded and thus can only be used to distinguish the presence or absence of a protein family. Proteins with a combined percentage of less than 1% across all 12 GELFrEE fractions were excluded from analyses.

To limit the complexity of the sample analysis while providing meaningful graphical trends, common names for protein families are utilized for grouping homologous protein

sequences. Protein families are determined on the basis of a comparative parsimony analysis where proteins are clustered according to shared peptide identifications. Supplemental Table 1 is a summary of identified currently accepted Ara h peanut allergens (1A),⁴² nonallergen proteins (1B), and recognized isoforms for each protein class. While there is a relatively low number of proteins in the *A. hypogaea* proteome as compared to mammalian and bacterial species, the presence of numerous isoforms in each protein family greatly complicates the proteomic investigation of the peanut proteome. This will be elaborated in greater detail below for the peanut allergen Ara h 3/4.

For the GELFrEE fractionated raw peanut lysate, all known peanut allergen classes were successfully identified and together comprise approximately 85% of the total protein content, with Ara h 1, Ara h 2, and Ara h 3/4 accounting for a combined 75%. High protein sequence coverage of peanut allergen proteins combined with the mass-selective fractionation of GELFrEE-LC-MS/MS provides a methodology for tracking peanut allergen peptides between tunable molecular weight boundaries. One of the most remarkable results of these experiments is the demonstration of the capability to identify unexpected trends in protein location on a gel. Comparisons of intact protein molecular weight to the estimated GELFrEE mass fractionation range can be used to pinpoint proteins deviating from the theoretical molecular weight (gel band location). Using this methodology, it was noted that some proteins were

detected in a majority of, if not all, GELFrEE fractions, in particular Conarachin (Q647H1_ARAHY), Ara h 1, Ara h 2, Ara h 3/4, and Ara h 6. While higher molecular weight proteins (Conarachin, Ara h1, and Ara h 3/4) were detected in multiple fractions as small protein cleavage fragments, detection of peanut 2S albumin (Ara h 2 and Ara h 6) peptides in fractions corresponding to molecular weights significantly greater (fractions 6–12; 33–98 kDa) than those of the intact protein (fractions 1–3; 0–23 kDa) was unexpected. This phenomenon, discussed in greater detail below, demonstrates a significant advantage of using a combined platform of in-solution GELFrEE-PAGE-LC–MS/MS to isolate unique components of protein structural rearrangements, interactions, and/or reactions that yield high molecular weight entities.

Improved Protein Coverage by Size-Selective GELFrEE Fractionation

To compare the information obtained from LC–MS/MS analysis of GELFrEE fractions with that of a more traditional whole peanut lysate digest, unfractionated sample extracts were evaluated by a bottom-up whole protein analysis. GELFrEE-LC–MS/MS and whole lysate LC–MS/MS sample systems were compared on the basis of the number of identified unique peptides corresponding to *A. hypogaea*. A Venn diagram (Figure 3A) summarizes the results between the GELFrEE-LC–MS/MS methodology (554 unique peptide identifications

among all twelve fractions) and a bottom-up whole lysate LC–MS/MS digest (296 unique peptides). Of the peptides identified, 268 peptides were common in both experimental platforms. Taking a more in-depth view of peptide identifications, Figure 3B presents a histogram comparing the number of unique peptide identifications between GELFrEE and whole lysate LC–MS/MS analyses according to peanut allergen protein families. For almost all of the protein allergens monitored, the GELFrEE method yielded a larger number of unique peptides. The most considerable difference in the number of peptides identified between the two methods is shown for the Ara h 3/4 protein group.

One explanation for the preferential recovery of Ara h 3/4 peptides in GELFrEE experiments likely results from the inability to distinguish isoforms when homologous tryptic peptides occur within a limited LC separation.^{43,44} Due to a large number of homologous isoforms, peptides shared between protein forms will be more abundant and more frequently detected, whereas sequences unique to a given isoform are present at lower abundances. Consequently, isoform distinctions are often lost or difficult to decipher by unfractionated methodologies as a result of separation inefficiencies and limited sampling rates. Thus, while a whole lysate approach is typically the model for a global proteome discovery-based experiment, only a fraction of the total peptide population for a given protein is identified, limiting knowledge of complex protein chemistries notoriously present in food samples.

The added degree of protein separation provided by GELFrEE fractionation not only improves protein resolution contributing to an overall enhancement in protein coverage, but also identifies peptides corresponding to isoform variants, missed cleavage products, and modified proteins. As shown in Figure 2, many proteins elicit peptide responses in several GELFrEE fractions, indicating the possibility of structural rearrangements or alternative protein forms that produce unique peptide identifications. Using the multidimensional platform of GELFrEE-LC–MS/MS, we can begin to understand some of the complex chemistries involving the peanut proteome and further contribute to the information database, from which more robust methods for peanut allergen detection may be developed. The applications of this sophisticated methodology are most clearly demonstrated in the investigation of isoform specificity, post-translational processing, and covalent complexation identified in peanut allergens Ara h 3/4 and Ara h 2 as discussed below.

Confirmation of Disulfide Complexes in Peanut Allergen Ara h 3/4

In combination with the progress in immunochemistry and genetic-based allergen detection methodologies, mass spectrometry provides a molecularly specific platform for differentiating between protein allergens and understanding protein chemistries in complex food systems. Utilizing the fractionation capabilities of GELFrEE, Figure 4 exemplifies the multidimensional approach of unreduced/reduced GELFrEE-PAGE for identifying disulfide-linked subunits. Figure 4A reproduces the one-dimensional SDS-PAGE of unreduced purified Ara h 1 (lane 2), purified Ara h 2 (lane 2), and a complete PBS extraction of raw peanut (lane 3). Sample preparations were performed without and with sample reduction to visualize disulfide-linked proteins based on differential band migration. While Ara h 1, an approximately 70 kDa protein, maintains

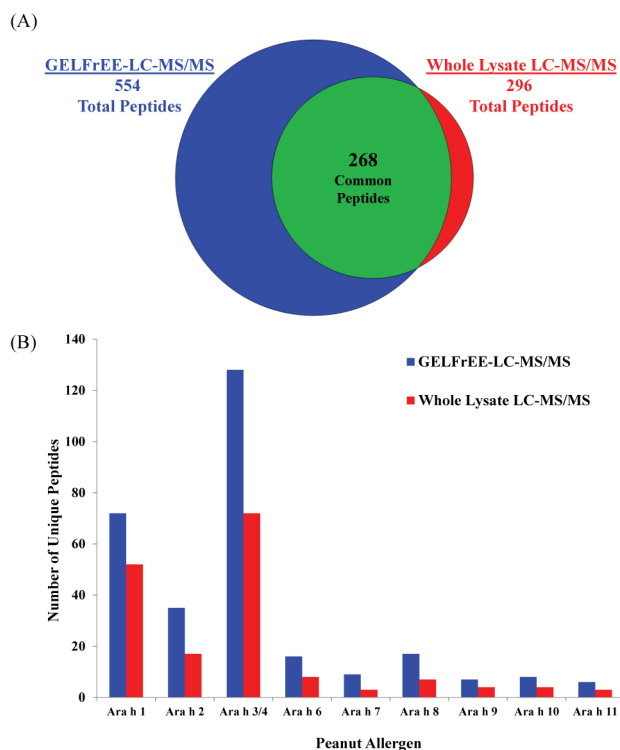


Figure 3. Comparative analysis of GELFrEE-LC–MS/MS versus unfractionated bottom-up LC–MS/MS analyses of raw peanut extract. (A) Venn diagram depicting the number of unique peptides identified using the GELFrEE-LC–MS/MS protein fractionation platform and a whole lysate LC–MS/MS bottom-up proteomics workflow. Total numbers of unique peptides are reported for each experimental method with common identifications shown in green. (B) Histogram plot of the number of unique peptides identified corresponding to specific allergen proteins using the GELFrEE-LC–MS/MS (blue) and whole lysate LC–MS/MS (red) platforms.

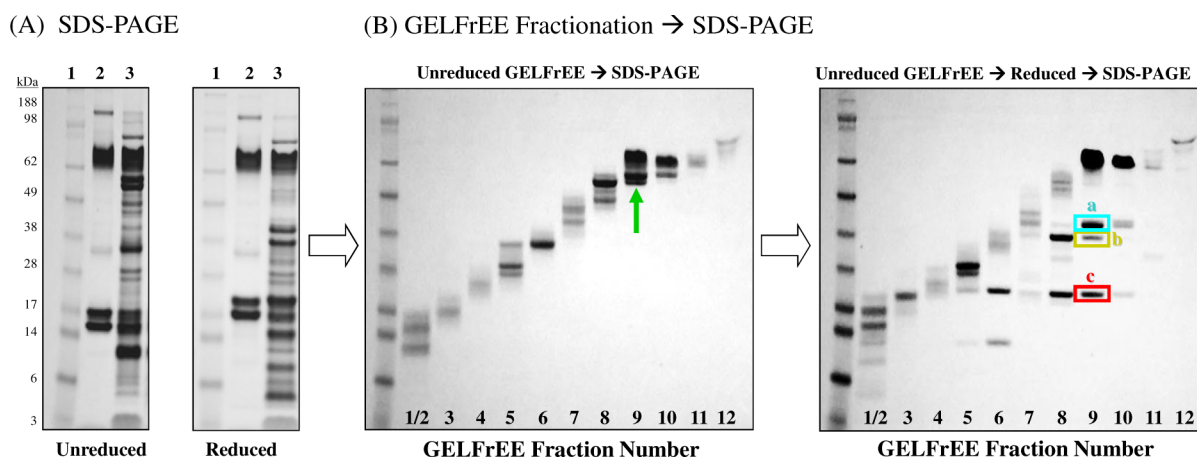


Figure 4. Visual identification of disulfide-linked proteins by GELFrEE-PAGE. (A) One-dimensional SDS-PAGE gels of unreduced and reduced whole peanut extracts. Lane 1: SeeBlue Plus2 prestained molecular weight ladder; Lane 2: purified peanut allergen standards Ara h 1 (~70 kDa) and Ara h 2.01/2.02 (16–18 kDa); Lane 3: PBS extract of raw peanut. (B) One-dimensional SDS-PAGE gel of unreduced GELFrEE fractions from PBS extracted raw peanuts. A second complementary SDS-PAGE gel illustrates the multidimensional approach of identifying disulfide-linked components by reducing individual GELFrEE fractions prior to gel-electrophoresis. Comparing banding patterns between unreduced and reduced sample conditions, protein bands in boxed regions “a”, “b”, and “c” of fraction 9 are visually resolved as potential disulfide-associated subunits originating from an unreduced molecular weight of approximately 60 kDa (green arrow). A SeeBlue Plus2 prestained protein ladder is provided for molecular weight reference.

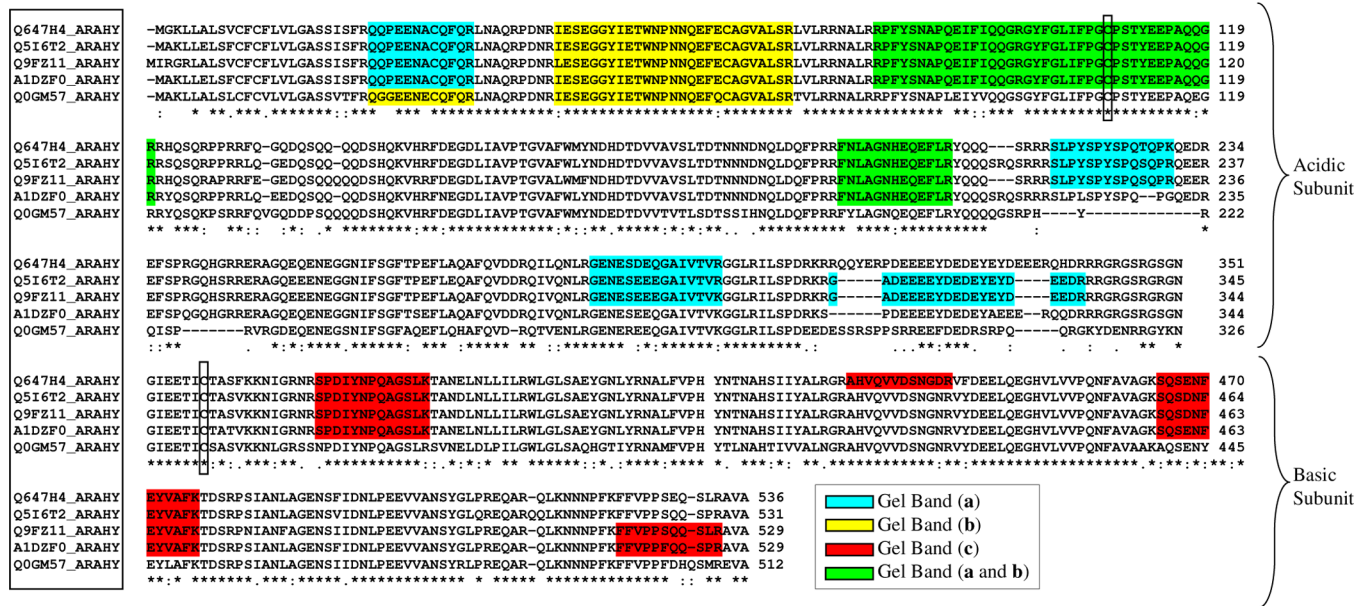


Figure 5. Sequence-based analysis of disulfide-associated Ara h 3/4 subunits as identified by GELFrEE-LC-MS/MS. Sequence alignments for a subset of homologous Ara h 3/4 proteins identified from select gel band digestions and LC-MS/MS analysis of protein bands “a”, “b”, and “c” in Figure 4B. Sequences were aligned using the ClustalW algorithm.⁵⁰ Peptide sequences are highlighted with a color that corresponds to the gel band (Figure 4B) from which they were parsimoniously identified. Aligned sequences are divided at the asparaginyl endopeptidase cleavage site, and the interchain cysteines involved in disulfide bonding are displayed along with the notation for the N-terminus acidic and C-terminus basic subunits. While complete isoform discrimination is difficult using a low resolution gel-based separation, mass spectrometric evaluation concluded peptides identified in gel bands “a” and “b” originate from the acidic domain of Ara h 3/4, and peptides identified in gel band “c” are specific to the basic chain of Ara h 3/4. Sequence notation is as follows: asterisk (*) corresponds to amino acid identity, colon (:) is a conserved substitution, and period (.) is a homologous substitution.

consistent migration between both gels, Ara h 2 isoforms appear as a doublet that migrates to a higher effective molecular weight when reduced (lane 2). This interesting phenomenon has been reported for other heavily disulfide-bonded proteins but is not well understood.⁴⁵ It is possible that the reduction of the four intramolecular disulfide bonds in Ara h 2 creates a less compact structure that alters the way the protein interacts with the gel matrix, producing a protein migration that more closely

matches the nominally accepted molecular weight (16–18 kDa). A similar trend is also seen in the whole peanut extract (lane 3); however, the complexity of an unfractionated peanut lysate makes it difficult to track the shifting of individual protein bands upon reduction of disulfide bonds.

To more clearly visualize differences in protein banding patterns, a whole protein PBS extract of raw peanut was analyzed by one-dimensional SDS-PAGE after mass-selective

fractionation by GELFrEE. Figure 4B illustrates the stepwise increase in molecular weight as unreduced peanut protein components are electrophoretically eluted from a gel cartridge. Providing a second dimension of separation, protein disulfide bonds are reduced after GELFrEE mass fractionation, but prior to SDS-PAGE. The added dimension of separation enables visual discrimination of protein fractions containing disulfide-linked subunits that separate into respective protein components. This phenomenon is most clearly demonstrated by comparing a single fraction between the unreduced and reduced sample conditions (Figure 4B). In an unreduced form, fraction 9 yields two protein bands at approximately 70 kDa and 60 kDa (green arrow). When raw peanut GELFrEE fractions are reduced prior to second dimension gel separation, fraction 9 reveals an unaltered 70 kDa protein band and three prominent lower molecular weight bands at approximate mass values of 38, 37, and 20 kDa as shown in respective boxed regions “a”, “b”, and “c”. Originating from an intact protein molecular weight of approximately 60 kDa, protein bands “a”, “b”, and “c” are visually resolved and targeted as potential disulfide-linked subunits. As compared to a traditional whole lysate analysis method in which proteins are typically reduced prior to digestion, thereby dissociating disulfide-linked complexes, this methodology of comparing unreduced and reduced protein GELFrEE fractions provides a mass-selective strategy for visualization, separation, and identification of such subunits.

To associate protein banding patterns with confirmed sequence information and validate the suspected disulfide-linked protein subunits, target protein bands “a”, “b”, and “c” were further investigated using in-gel digestions and mass spectrometry (Figure 5). Tryptic LC-MS/MS digestions of gel bands “a”, “b”, and “c” yielded identification of peptides that correspond to Ara h 3/4 in all three bands. The protein identity of the selected protein bands agrees well with findings in the literature that report the protein allergen Ara h 3/4 is post-translationally cleaved by an asparaginyl endopeptidase into a disulfide-linked acidic and basic subunit.^{12–14} In an unreduced form, the two subunits remain covalently linked by an intermolecular disulfide bridge with a nominal molecular weight of 60 kDa; however, after sample reduction, the protein separates into respective isolated subunits.

The sequence alignment shown in Figure 5 represents a limited protein parsimony report displaying only isoform-specific peptides for Ara h 3/4. Mass spectrometric evaluation concluded that peptides identified in gel bands “a” and “b” originate from the acidic domain of the protein, whereas peptides identified in gel band “c” are specific to the basic chain of Ara h 3/4. Both representing peptide components in the acidic chain of Ara h 3/4, the question arises on whether protein bands “a” and “b” vary in isoform content or experience differential band migration as a result of variable proteolytic processing. Due to inefficiencies in the resolving capabilities of protein gel electrophoresis, multiple Ara h 3/4 homologous protein sequences are identified in each isolated gel band. Comparing peptide recovery between protein bands “a” and “b” in Figure 5, however, reveals the majority of the N-terminal region of Ara h 3/4 to encompass band “a”, while protein band “b” reports peptide coverage in only the first two-thirds of the acidic domain. The apparent C-terminal truncation of the acidic subunit, thus, can be used to explain differences in protein migration of isolated gel bands “a” and “b”.

Acidic domain truncation by post-translational proteolytic processing can also be evaluated by an anti-Ara h 3 Western blot (Supplemental Figure 1) under reduced GELFrEE sample conditions. The Western blot reveals the presence of less abundant Ara h 3/4 polypeptide fragments in addition to isolated protein bands at 38 kDa, 37 kDa, and 20 kDa. A more thorough gel cutting and LC-MS/MS analysis was performed on the majority of gel bands whose migration patterns emerged in the reduced SDS-PAGE (data not shown). Utilizing Western blotting as a survey to identify the gel migration of Ara h 3 and mass spectrometry to provide molecular sequence information, acidic subunit proteolytic peptides were identified as the primary component for many of the isolated gel bands in molecular weight ranges between 14 and 38 kDa. Relating back to findings from Figure 2, extensive post-translational processing of Ara h 3/4 is further supported by the identification of protein components in all GELFrEE fractions.

Using mass spectrometry to identify and characterize Ara h 3/4 polypeptide fragments is complicated by the presence of numerous homologous sequences contained within the *A. hypogaea* protein database. Proteins that share peptide sequences can be grouped into clusters to aid in proteomic analyses. In GELFrEE-LC-MS/MS analyses, 15 isoforms are identified in the Ara h 3/4 protein cluster with an overall 52% homology between all arachin isoforms and sequence fragments. Using general guidelines of protein parsimony, only 9 of the 15 isoforms are identified by unique peptides in all GELFrEE fractions. Protein alignment and classification are presented in Supplemental Figures 2 and 3 in attempts to group known protein variants into three major classifications based upon sequence homology. Understanding similarities in homologous protein sequences aids to assign indeterminate peptides and minimize redundancy in identifying protein sequences for individual protein bands.

As illustrated with Ara h 3/4, excessive post-translational processing combined with overlapping homologous protein isoforms demonstrate the unexpected complexity associated with peanut proteomics. A multidimensional fractionation platform helps reduce sample complexities, allowing a deeper investigation of underlying protein chemistry. Visual identification of disulfide-linked protein complexes, as shown in the analysis of Ara h 3/4, demonstrates the power of unreduced/reduced mass fractionation to isolate protein components without the need for purified protein. Unlike other low resolution and labor intensive methods for examining protein disulfide-bond formation, such as redox two-dimensional SDS-PAGE,^{46,47} the GELFrEE-PAGE-LC-MS/MS platform can be utilized to confirm disulfide-associated protein complexes and isolate solution-phase fractions for downstream characterization and molecular authentication by LC-MS/MS in a complex matrix.

Evidence for Structural Changes in *Arachis hypogaea* 2S Albumins

As noted previously, several protein allergens or allergen fragments are identified by mass spectrometry in multiple GELFrEE fractions other than the fraction(s) corresponding to the intact molecular weight of the protein. Although it is not unexpected that fragments of higher molecular weight proteins are also present as truncated protein forms resulting from proteolytic processing (as in the case of Ara h 1 and Ara h 3/4),²⁸ Ara h 2 represents a case in which a low molecular weight protein appears in higher molecular weight fractions. This

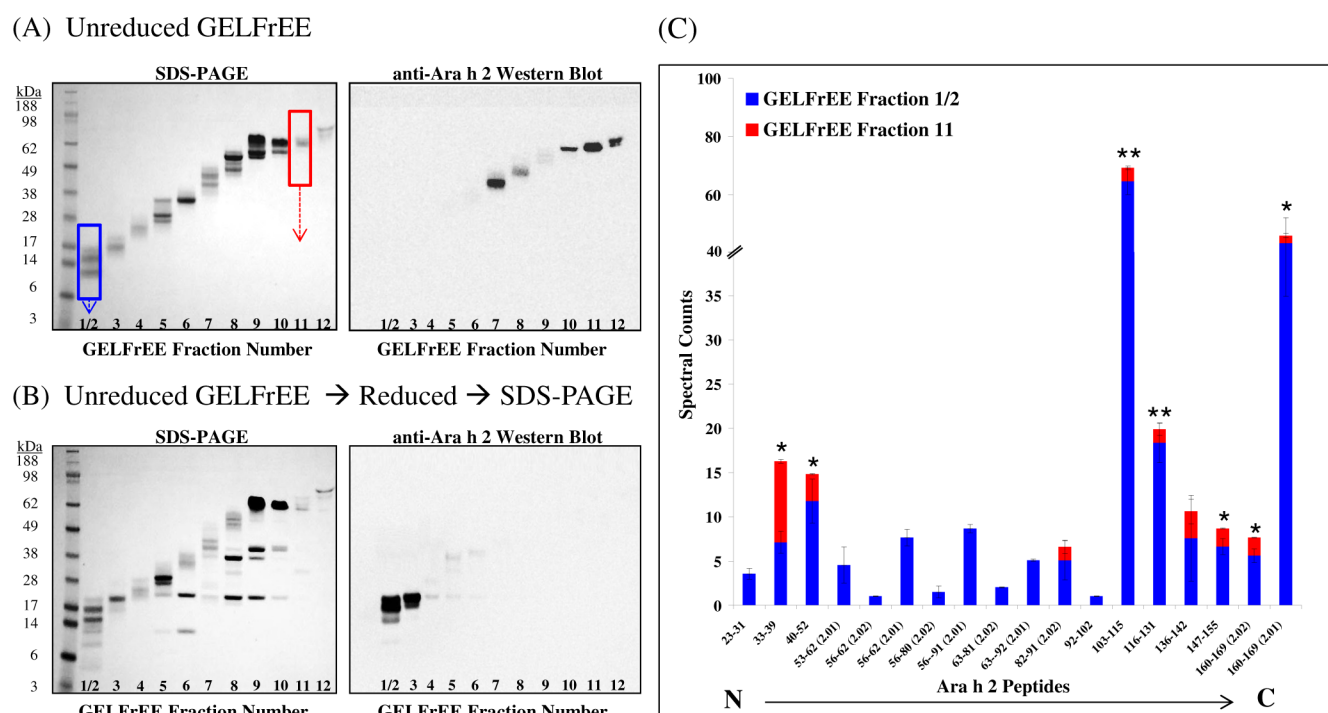


Figure 6. Evidence for high molecular weight associations of *Arachis hypogaea* 2S albumins. One-dimensional SDS-PAGE gels and anti-Ara h 2 Western blots for unreduced (A) and reduced (B) raw peanut PBS extracts fractionated using GELFrEE. Protein bands for Western blots were visualized using enhanced chemiluminescence imaging. (C) Relative quantification by spectral counting for all Ara h 2 peptides identified by mass spectrometric analysis of GELFrEE fractions 1/2 and 11. Peptides are ordered from N- to C-terminus of Ara h 2 and recorded by residue number. Isoform classifications are provided for Ara h 2.01 (P2/P4) and Ara h 2.02 (P1/P3). The presence of a double dash (--) signifies a break in sequential residue notation and the absence of the 12-residue insert for the Ara h 2.01 isoform pair. Spectral count values are summed for fully tryptic and associated missed cleavage peptides. Peptides are signified by an asterisk (*) to indicate the number of cysteine residues in the identified peptide.

phenomenon is particularly intriguing as the multidimensional separation of GELFrEE-PAGE fractionation should degrade all but covalent interactions.

The identification of Ara h 2 as a participant in higher molecular weight covalent interactions can be further supported by Western blot analyses. Figure 6A and B represents PAGE gels and anti-Ara h 2 Western blots of unreduced and reduced GELFrEE samples, respectively. SDS-PAGE gels (replicated from Figure 4B) are presented as a reference to align protein banding patterns with corresponding Western blots. For the unreduced GELFrEE separation, the anti-Ara h 2 Western blot (Figure 6A) provides evidence of Ara h 2 (antibody epitopes) in high molecular weight GELFrEE fractions 6–12. Unexpectedly, however, antibody interaction with the native protein (nominal molecular weight of 16–18 kDa) is not observed for unreduced sample conditions. Reduction of protein disulfide bonds in GELFrEE fractions (Figure 6B), contrastingly, results in a conformational change exposing available epitopes for antibody recognition producing a protein doublet at the predicted molecular weight range between 16 and 18 kDa (lanes 1/2 and 3). Similar results in protein recognition were reported for alternative Ara h 2 antibodies (data not shown). From these results, it may be hypothesized that ambiguous antibody-protein recognition between unreduced and reduced sample conditions is dependent on disulfide bonding (i.e., tertiary structure) and therefore cysteine-containing domains of Ara h 2. In contrast, the anti-Ara h 1 antibody consistently recognizes Ara h 1 (no reported disulfide bonds) in unreduced and reduced forms (Supplemental Figure 4).

While informative as a qualitative measure of protein identification, Western blot analysis is subject to the vagaries

of the protein structure (linear or conformational epitopes) to which the antibodies are raised. Such information regarding antibody-protein recognition, however, is often unknown. To supplement our previous findings, purified Ara h 2 was prepared and analyzed using the same GELFrEE method as that used with the lysate (data not shown). Results observed for the purified protein yielded antibody recognition only in the expected molecular weight gel locations (fraction 1/2 and 3) for the reduced form of the protein. Furthermore, the absence of Ara h 2 protein recognition in higher molecular weight fractions for the purified standard suggests Ara h 2 protein complexes are endogenous to the raw peanut. While immunochemical recognition is the standard for allergen detection within the food community, ambiguous antibody-protein interactions limit the ability to decode protein behavior. To overcome the limitations of antibody-based detection, mass spectrometry was used to perform a structurally unbiased sample analysis, whereby modified proteins and protein fragments can be identified and confirmed.

Molecular distributions of Ara h 2 peptides were monitored in GELFrEE fractions corresponding to the native Ara h 2 mass (fraction 1/2) and a higher molecular weight segment (fraction 11). Figure 6C displays the relative abundance (by spectral counts) of individual Ara h 2 peptides plotted as a function of aligned peptide residue number ordered from the N- to C-terminus of the protein. It is evident from this figure that the majority of peptides identified in the higher molecular weight GELFrEE fraction are characterized by cysteine-containing sequences (*). Furthermore, changes in Ara h 2 peptide distributions are noted between the two gel bands excised from greatly different molecular weight locations. Although we are

unable to declare with any degree of certainty, it appears as though the Ara h 2 peptides in fraction 11 are represented by cysteine-containing peptide fragments involved in higher molecular weight protein complexes.

According to the crystal structure reported for Ara h 2, the protein is composed of five helices and four disulfide bonds with at least one disulfide bond connecting each helix to one another.⁴⁸ The flexible loop between helices two and three maintains a disordered structure making the 2.02 isoform sequences sensitive to proteolysis. While there was no evidence of proteolysis in the 2010 report by Mueller et al.,⁴⁸ the potential for disulfide rearrangement may result in protein restructuring and formation of higher molecular weight protein complexes. To circumvent the likelihood of non-native disulfide scrambling, sample extracts were alkylated with iodoacetamide prior to GELFrEE fractionation.

Utilizing the benefits of the GELFrEE-PAGE-LC-MS/MS platform, we were able to provide peptide-specific information that suggests Ara h 2 undergoes structural rearrangements that yield higher molecular weight cysteine-containing entities. While the exact conformation and functional role of Ara h 2-containing entities in high mass GELFrEE fractions are yet to be determined, this finding provides evidence of heretofore unknown Ara h 2 protein forms or rearrangements in raw peanut. Other members of the 2S albumin family in *A. hypogaea* (Ara h 6 and Ara h 7) exhibit a high degree of sequence homology with a characteristic pattern of disulfide linkages as shown in the aligned sequences presented in Supplemental Figure 5. While less abundant in total protein content as compared to Ara h 2, mass spectrometry results provided similar evidence of higher order protein rearrangements for Ara h 6 and to a lesser extent Ara h 7. Attempts to understand the role of protein tertiary structure in molecular mechanisms of an allergic reaction is crucial for the development of high-throughput, robust methods for allergen detection and quantification in food. Furthermore, structural features including disulfide bonds and protein functionality are suggested as predictive measures to aid the development of immunotherapy and potential strategies to alleviate allergen response.⁴⁹

CONCLUSIONS

Peanuts rank among the most allergenic foods. For an allergic individual, total avoidance of the allergen-containing food is currently the only effective treatment. Despite patient vigilance, allergy-affected individuals must remain continually conscious of the widespread use of peanuts as food ingredients and of the possibility of peanut contamination in processed foods. As a result, reliable detection and quantification methods for food allergens are necessary in order to ensure food labeling policy compliance and improve consumer protection.

With a rise in the number of patients diagnosed with food allergy over the past 10 years, a shift toward mass spectrometry based strategies for food allergen detection has proven to be an essential advancement in food science and technology. A challenge in the development of methods for allergen detection is an understanding of the protein structure and the functional relationships of protein conformation and allergenicity. Toward this goal, we demonstrated here the necessity and advantage of a multidimensional global proteome discovery study of *A. hypogaea*.

The protein fractionation capabilities of a GELFrEE-LC-MS/MS proteomics strategy combined with the visual

validation of protein migration by SDS-PAGE (stained or Western blot analysis) constitute a multidimensional platform enabling rigorous interrogation of the peanut proteome. The unique capability of this combined analytical method allows for the identification of low abundance proteins while providing supporting information regarding isoform discrimination, post-translational proteolytic processing, and covalent modifications that are often overlooked or inaccessible in a direct bottom-up proteomics approach. Furthermore, mass selective fractions can be compared under unreduced and reduced sample conditions and visualized by SDS-PAGE to allow identification of potential disulfide-linked protein complexes. The extra dimension of separation complemented with molecularly specific mass spectrometric information provides a means to characterize some of the complex chemistry within the peanut proteome. One of the most interesting findings uniquely provided by the GELFrEE-PAGE-LC-MS/MS platform is the unforeseen identification of peanut allergen protein components in molecular weight fractions much higher than the nominal molecular weight of the protein. This was most notably displayed for Ara h 2 but also was shown to a lesser extent for other 2S albumins in *A. hypogaea*. From GELFrEE-LC-MS/MS experiments, we can conclude that many of the peanut protein allergens demonstrate evidence of protein components existing in a variety of different conformations over a wide range of molecular weight entities. Understanding the chemistry behind protein interactions, rearrangements, and modifications will further aid the development of quantitative methods for peanut detection in a variety of complex foods.

ASSOCIATED CONTENT

Supporting Information

Table 1 and Supplemental Figures 1–5. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Ph: (240) 402-2019. Fax: (301) 436-2624. E-mail: Christine.Hebling@fda.hhs.gov.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We gratefully acknowledge Bruce Kotz of the Golden Peanut Company in Alpharetta, Georgia for providing the Raw Runner peanut reference samples. Our kind regards go to Jinxi Li and Kevin Shefcheck for their early work in the peanut allergen project during their tenure in our group at the FDA and to Sara McGrath for useful discussions. We also thank Dr. Soheila Maleki at the U.S. Department of Agriculture for providing Ara h 1, Ara h 2, and Ara h 3 purified protein and antibodies. The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. Food and Drug Administration.

REFERENCES

- (1) Sicherer, S. H.; Sampson, H. A. Peanut allergy: emerging concepts and approaches for an apparent epidemic. *J. Allergy Clin. Immunol.* **2007**, *120* (3), 491–503.
- (2) Akagawa, M.; Handoyo, T.; Ishii, T.; Kumazawa, S.; Morita, N.; Suyama, K. Proteomic analysis of wheat flour allergens. *J. Agric. Food Chem.* **2007**, *55* (17), 6863–6870.

- (3) Picariello, G.; Mamone, G.; Addeo, F.; Ferranti, P. The frontiers of mass spectrometry-based techniques in food allergenomics. *J. Chromatogr. A* **2011**, *1218* (42), 7386–7398.
- (4) Shin, D. S.; Compadre, C. M.; Maleki, S. J.; Kopper, R. A.; Sampson, H.; Huang, S. K.; Burks, A. W.; Bannion, G. A. Biochemical and structural analysis of the IgE binding sites on Ara h 1, an abundant and highly allergenic peanut protein. *J. Biol. Chem.* **1998**, *273* (22), 13753–13759.
- (5) Koppelman, S.; Wensing, M.; Ertmann, M.; Knulst, A.; Knol, E. Relevance of Ara h1, Ara h 2 and Ara h 3 in peanut-allergic patients, as determined by immunoglobulin E Western blotting, basophil-histamine release and intracutaneous testing: Ara h 2 is the most important peanut allergen. *Clin. Exp. Allergy* **2004**, *34* (4), 583–590.
- (6) Palmer, G.; Dibbern, D.; Burks, A.; Bannion, G.; Bock, S.; Porterfield, H.; McDermott, R.; Dreskin, S. Comparative potency of Ara h 1 and Ara h 2 in immunochemical and functional assays of allergenicity. *Clin. Immunol.* **2005**, *115* (3), 302–312.
- (7) van Bortel, E. L.; van Beers, M. M.; Koppelman, S. J.; van den Broek, L. A.; Gruppen, H. Allergen Ara h 1 occurs in peanuts as a large oligomer rather than as a trimer. *J. Agric. Food Chem.* **2006**, *54* (19), 7180–7186.
- (8) Adachi, M.; Kanamori, J.; Masuda, T.; Yagasaki, K.; Kitamura, K.; Mikami, B.; Utsumi, S. Crystal structure of soybean 11S globulin: Glycinin A3B4 homohexamers. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (12), 7395–7400.
- (9) Boldt, A.; Fortunato, D.; Conti, A.; Petersen, A.; Ballmer-Weber, B.; Lepp, U.; Reese, G.; Becker, W. Analysis of the composition of an immunoglobulin E reactive high molecular weight protein complex of peanut extract containing Ara h 1 and Ara h 3/4. *Proteomics* **2005**, *5* (3), 675–686.
- (10) Guo, B. Z.; Liang, X. Q.; Chung, S. Y.; Holbrook, C. C.; Maleki, S. J. Proteomic analysis of peanut seed storage proteins and genetic variation in a potential peanut allergen. *Protein Peptide Lett.* **2008**, *15* (6), 567–577.
- (11) Liang, X. Q.; Luo, M.; Holbrook, C. C.; Guo, B. Z. Storage protein profiles in Spanish and runner market type peanuts and potential markers. *BMC Plant Biol.* **2006**, *6*, 24.
- (12) Dickinson, C.; Hussein, E.; Nielsen, N. Role of posttranslational cleavage in glycinin assembly. *Plant Cell* **1989**, *1* (4), 459–469.
- (13) Jung, R.; Scott, M.; Nam, Y.; Beaman, T.; Bassuner, R.; Saalbach, I.; Muntz, K.; Nielsen, N. The role of proteolysis in the processing and assembly of 11S seed globulins. *Plant Cell* **1998**, *10* (3), 343–357.
- (14) Scott, M.; Jung, R.; Muntz, K.; Nielsen, N. A protease responsible for post-translational cleavage of a conserved Asn-Gly linkage in glycinin, the major seed storage protein of soybean. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89* (2), 658–662.
- (15) Koppelman, S. J.; Knol, E. F.; Vlooswijk, R. A.; Wensing, M.; Knulst, A. C.; Hefle, S. L.; Gruppen, H.; Piersma, S. Peanut allergen Ara h 3: isolation from peanuts and biochemical characterization. *Allergy* **2003**, *58* (11), 1144–1151.
- (16) Piersma, S. R.; Gaspari, M.; Hefle, S. L.; Koppelman, S. J. Proteolytic processing of the peanut allergen Ara h 3. *Mol. Nutr. Food Res.* **2005**, *49* (8), 744–755.
- (17) Stanley, J.; King, N.; Burks, A.; Huang, S.; Sampson, H.; Cockrell, G.; Helm, R.; West, C.; Bannion, G. Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. *Arch. Biochem. Biophys.* **1997**, *342* (2), 244–253.
- (18) Burks, A.; Williams, L.; Connaughton, C.; Cockrell, G.; O'Brien, T.; Helm, R. Identification and characterization of a second major peanut allergen, Ara h II, with use of the sera of patients with atopic dermatitis and positive peanut challenge. *J. Allergy Clin. Immunol.* **1992**, *90* (6), 962–969.
- (19) Chatel, J. M.; Bernard, H.; Orson, F. M. Isolation and characterization of two complete Ara h 2 isoforms cDNA. *Int. Arch. Allergy Immunol.* **2003**, *131* (1), 14–18.
- (20) Li, J.; Shefcheck, K.; Callahan, J.; Fenselau, C. Primary sequence and site-selective hydroxylation of prolines in isoforms of a major peanut allergen protein Ara h 2. *Protein Sci.* **2010**, *19* (1), 174–182.
- (21) Cabanos, C.; Tandang-Silvas, M. R.; Odijk, V.; Brostedt, P.; Tanaka, A.; Utsumi, S.; Maruyama, N. Expression, purification, cross-reactivity and homology modeling of peanut profilin. *Protein Expression Purif.* **2010**, *73* (1), 36–45.
- (22) Mittag, D.; Akkerdaas, J.; Ballmer-Weber, B. K.; Vogel, L.; Wensing, M.; Becker, W. M.; Koppelman, S. J.; Knulst, A. C.; Helbling, A.; Hefle, S. L.; Van Ree, R.; Vieths, S. Ara h 8, a Bet v 1-homologous allergen from peanut, is a major allergen in patients with combined birch pollen and peanut allergy. *J. Allergy Clin. Immunol.* **2004**, *114* (6), 1410–1417.
- (23) Krause, S.; Reese, G.; Randow, S.; Zennaro, D.; Quarantino, D.; Palazzo, P.; Ciardiello, M.; Petersen, A.; Becker, W.; Mari, A. Lipid transfer protein (Ara h 9) as a new peanut allergen relevant for a Mediterranean allergic population. *J. Allergy Clin. Immunol.* **2009**, *124* (4), 771–778.
- (24) Lauer, I.; Dueringer, N.; Pokoj, S.; Rehm, S.; Zoccatelli, G.; Reese, G.; Miguel-Moncin, M.; Cistero-Bahima, A.; Enrique, E.; Lidholm, J.; Vieths, S.; Scheurer, S. The non-specific lipid transfer protein, Ara h 9, is an important allergen in peanut. *Clin. Exp. Allergy* **2009**, *39* (9), 1427–1437.
- (25) Pons, L.; Chery, C.; Romano, A.; Namour, F.; Artesani, M.; Gueant, J. The 18 kDa peanut oleosin is a candidate allergen for IgE-mediated reactions to peanuts. *Allergy* **2002**, *57*, 88–93.
- (26) Chassaigne, H.; Tregoeat, V.; Norgaard, J. V.; Maleki, S. J.; van Hengel, A. J. Resolution and identification of major peanut allergens using a combination of fluorescence two-dimensional differential gel electrophoresis, Western blotting and Q-TOF mass spectrometry. *J. Proteomics* **2009**, *72* (3), 511–526.
- (27) Kottapalli, K. R.; Payton, P.; Rakwal, R.; Agrawal, G. K.; Shibato, J.; Burrow, M.; Puppala, N. Proteomics analysis of mature seed of four peanut cultivars using two-dimensional gel electrophoresis reveals distinct differential expression of storage, anti-nutritional, and allergenic proteins. *Plant Sci.* **2008**, *175* (3), 321–329.
- (28) Schmidt, H.; Gelhaus, C.; Latendorf, T.; Nebendahl, M.; Petersen, A.; Krause, S.; Leippe, M.; Becker, W. M.; Janssen, O. 2-D DIGE analysis of the proteome of extracts from peanut variants reveals striking differences in major allergen contents. *Proteomics* **2009**, *9* (13), 3507–3521.
- (29) Jovin, T.; Chrambach, A.; Naughton, M. A. An apparatus for preparative temperature-regulated polyacrylamide gel electrophoresis. *Anal. Biochem.* **1964**, *9*, 351–369.
- (30) Lewis, U. J.; Clark, M. O. Preparative methods for disk electrophoresis with special reference to the isolation of pituitary hormones. *Anal. Biochem.* **1963**, *6*, 303–315.
- (31) Tran, J.; Doucette, A. Gel-eluted liquid fraction entrapment electrophoresis: An electrophoretic method for broad molecular weight range proteome separation. *Anal. Chem.* **2008**, *80* (5), 1568–1573.
- (32) Tran, J.; Doucette, A. Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Anal. Chem.* **2009**, *81* (15), 6201–6209.
- (33) Kellie, J.; Tran, J.; Lee, J.; Ahlf, D.; Thomas, H.; Ntai, I.; Catherman, A.; Durbin, K.; Zamdborg, L.; Vellaichamy, A.; Thomas, P.; Kelleher, N. The emerging process of top-down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput. *Mol. Biosyst.* **2010**, *6* (9), 1532–1539.
- (34) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **2011**, *480* (7376), 254–258.
- (35) Bora, A.; Anderson, C.; Bachani, M.; Nath, A.; Cotter, R. J. Robust two-dimensional separation of intact proteins for bottom-up tandem mass spectrometry of the human CSF proteome. *J. Proteome Res.* **2012**, *11* (6), 3143–3149.

- (36) Botelho, D.; Wall, M. J.; Vieira, D. B.; Fitzsimmons, S.; Liu, F.; Doucette, A. Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation. *J. Proteome Res.* **2010**, *9* (6), 2863–2870.
- (37) Schmitt, D. A.; Nesbit, J. B.; Hurlburt, B. K.; Cheng, H.; Maleki, S. J. Processing can alter the properties of peanut extract preparations. *J. Agric. Food Chem.* **2010**, *58* (2), 1138–1143.
- (38) Koppelman, S. J.; Bruijnzeel-Koomen, C. A.; Hessing, M.; de Jongh, H. H. Heat-induced conformational changes of Ara h 1, a major peanut allergen, do not affect its allergenic properties. *J. Biol. Chem.* **1999**, *274* (8), 4770–4777.
- (39) Sen, M.; Kopper, R.; Pons, L.; Abraham, E. C.; Burks, A. W.; Bannon, G. A. Protein structure plays a critical role in peanut allergen stability and may determine immunodominant IgE-binding epitopes. *J. Immunol.* **2002**, *169* (2), 882–887.
- (40) Slotta, D.; McFarland, M.; Markey, S. MassSieve: Panning MS/MS peptide data for proteins. *Proteomics* **2010**, *10* (16), 3035–3039.
- (41) Poms, R.; Klein, C.; Anklam, E. Methods for allergen analysis in food: a review. *Food Addit. Contam.* **2004**, *21* (1), 1–31.
- (42) Mari, A.; Scala, E.; Palazzo, P.; Ridolfi, S.; Zennaro, D.; Carabella, G. Bioinformatics applied to allergy: Allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. *Cell. Immunol.* **2006**, *244* (2), 97–100.
- (43) Duncan, M. W.; Aebersold, R.; Caprioli, R. M. The pros and cons of peptide-centric proteomics. *Nat. Biotechnol.* **2010**, *28* (7), 659–664.
- (44) Nesvizhskii, A. I.; Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* **2005**, *4* (10), 1419–1440.
- (45) Bardwell, J. C.; Lee, J. O.; Jander, G.; Martin, N.; Belin, D.; Beckwith, J. A pathway for disulfide bond formation in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90* (3), 1038–1042.
- (46) Cumming, R.; Andon, N.; Haynes, P.; Park, M.; Fischer, W.; Schubert, D. Protein disulfide bond formation in the cytoplasm during oxidative stress. *J. Biol. Chem.* **2004**, *279* (21), 21749–21758.
- (47) Yano, H.; Wong, J.; Lee, Y.; Cho, M.; Buchanan, B. A strategy for the identification of proteins targeted by thioredoxin. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (8), 4794–4799.
- (48) Mueller, G. A.; Gosavi, R. A.; Pomés, A.; Wünschmann, S.; Moon, A. F.; London, R. E.; Pedersen, L. C. Ara h 2: crystal structure and IgE binding distinguish two subpopulations of peanut allergic patients by epitope diversity. *Allergy* **2011**, *66* (7), 878–885.
- (49) Yano, H. Disulfide-related proteomic studies on food allergens. *Expert Rev. Proteomics* **2009**, *6* (5), S63–S71.
- (50) Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinf.* **2002**, Chapter 2, Unit 2.3.

Size-Selective Fractionation and Visual Mapping of Allergen Protein Chemistry in *Arachis hypogaea*

Christine M. Hebling, Mark M. Ross, John H. Callahan, and Melinda A. McFarland*

U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition,

5100 Paint Branch Parkway, College Park, MD 20740

Supporting Information

SUPPORTING INFORMATION TABLE AND FIGURE LEGENDS

Supplemental Table 1 – Compilation of identified protein allergens (A) and non-allergens (B) grouped into protein families that are obtained from the GELFrEE-LC-MS/MS and whole lysate LC-MS/MS digestion platforms. The parsimony type is indicated for each protein identified, with subset and subsumable proteins filtered from the parsimony report. For example, the Ara h 1 (7S Globulin) protein family is represented by the following SwissProt protein names: ALL11_ARAHY, ALL12_ARAHY, B3IXL2_ARAHY, Q6PSU3_ARAHY, Q6PSU4_ARAHY, Q6PSU5_ARAHY, and Q6PSU6_ARAHY. In this protein family, ALL11_ARAHY, B3IXL2_ARAHY and Q6PSU3_ARAHY are identified by equivalent peptides and ALL12_ARAHY, Q6PSU5_ARAHY, and Q6PSU6_ARAHY are differentiable by parsimony comparison for GELFrEE sample fractions analyzed. Protein sequences recognized as allergen standards in the allergen community are indicated (www.allergen.org).

Supplemental Figure 1 - Confirmation of disulfide-associated Ara h 3/4 protein subunits by GELFrEE-Western blot. (A) One-dimensional SDS-PAGE gel of reduced GELFrEE fractions from PBS-extracted raw peanut. (B) Western blot analyses using Chicken anti-Ara h 3 on reduced GELFrEE fractions indicate the presence of Ara h 3 in numerous gel bands. A SeeBlue Plus2 prestained protein ladder is provided for molecular weight reference.

Supplemental Figure 2 – Geneious Pro 5.5.6 created by Biomatters (available from <http://www.geneious.com/>) multiple sequence alignment for Ara h 3.0101 (O82580_ARAHY; Sequence 1), Ara h 3.0201 (Q9SQH7_ARAHY; Sequence 2), and eight additional protein isoforms (Sequences 3-10). Percent homology, or the percentage of columns in the alignment in which all residues are identical, is reported as 81% for all ten sequences. Global sequence alignment was accomplished using default cost matrix parameters. The consensus sequence is provided for comparison with the aligned protein sequences. Single letter amino acid notation is provided for the consensus sequence for identical residues. Ambiguous amino acids, asparagine/aspartic acid, glutamine/glutamic acid, and leucine/isoleucine are represented by single letter residues B, Z, or J, respectively. Conserved residues matching by identity or homology are shaded in yellow and variable regions in white. Gaps are noted with a dashed line or red bar.

Supplemental Figure 3 – Geneious Pro 5.5.6 created by Biomatters (available from <http://www.geneious.com/>) multiple sequence alignment for Ara h 3.0101 (O82580_ARAHY; Sequence 1), Ara h 3.0201 (Q9SQH7_ARAHY; Sequence 2), and two sub-classes of isoform alignments (A and B). Arachin isoforms (A), Q6IWG5_ARAHY, Q0GM57_ARAHY, and AHY3_ARAHY (Sequences 3, 4, and 5 respectively), share a 79% homology with one another and 57% homology as compared to Ara h 3 standard sequences. The Arachin sub-classification is shown to contain extended gap regions within the acidic chain leading to regions of variability within the combined sequence alignment. Panel B, includes the A1DZF1_ARAHY and Q8LL03_ARAHY protein sequences associated with the common names of Arachin 7 and Trypsin Inhibitor, respectively (Sequences 3 and 4). These protein fragments are closely related to the N-terminal domain of Ara h 3 (Sequences 1 and 2), with an 89% homology. Global sequence alignment is completed using default cost matrix parameters. The consensus sequence is provided for comparison with the aligned protein sequences. Single letter amino acid notation is provided for the consensus sequence for identical residues. Ambiguous amino acids, asparagine/aspartic acid, glutamine/glutamic acid, and leucine/isoleucine are represented by single letter residues B, Z, or J, respectively. Conserved residues matching by identity or homology are shaded in yellow and variable regions in white. Gaps are noted with a dashed line or red bar.

Supplemental Figure 4 – One-dimensional SDS-PAGE gels (1, 3) and anti-Ara h 1 Western blots (2, 4) for unreduced (1, 2) and reduced (3, 4) raw peanut PBS extracts fractionated using the GELFrEE system. Protein bands from Western blots were visualized using enhanced chemiluminescence imaging. A SeeBlue Plus2 prestained protein ladder is provided for molecular weight reference.

Supplemental Figure 5 – Geneious Pro 5.5.6 created by Biomatters (available from <http://www.geneious.com/>) multiple sequence alignment for Ara h 2 (CONG7_ARAHY; 2.01-P2/P4 and 2.02-P3/P1), Ara h 6 (CONG_ARAHY), Ara h 7.0101 (Q9SQH1_ARAHY), and Ara h 7.0201 (B4XID4_ARAHY). Percent homology, or the percentage of columns in the alignment in which all residues are identical, is reported as 25%. Global sequence alignment is completed using default cost matrix parameters. The consensus sequence is provided for comparison with the aligned protein sequences. Single letter amino acid notation is provided for the consensus sequence for identical residues. Ambiguous amino acids, asparagine/aspartic acid, glutamine/glutamic acid, and leucine/isoleucine are represented by single letter residues B, Z, or J, respectively. Conserved residues matching by identity or homology are shaded in yellow and variable regions in white. Gaps are noted with a dashed line or red bar.

(A) *Arachis hypogaea* Protein Allergens Identified in GELFrEE and Whole Protein Lysate Digestions

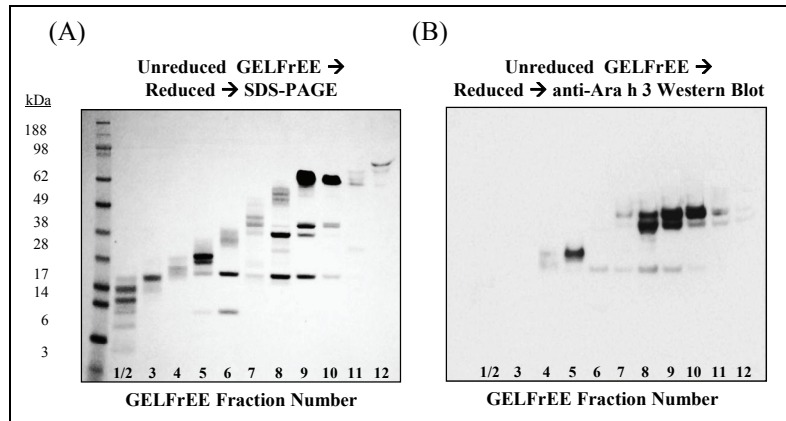
Protein Name	Mass	Protein Family	IUIS	GELFrEE FRACTIONATION			LYSATE DIGESTION		
				Parsimony Type	Peptides	% Coverage	Parsimony Type	Peptides	% Coverage
ALL11_ARAHY	70639.4	Ara h 1 (7S Globulin)	ARA H 1.0101	EQUIVALENT	52	59.3	EQUIVALENT	41	53.7
ALL12_ARAHY	71700.9	Ara h 1 (7S Globulin)		DIFFERENTIABLE	51	61.5	DIFFERENTIABLE	43	57.5
B3IXL2_ARAHY	70639.4	Ara h 1 (7S Globulin)		EQUIVALENT	52	59.3	EQUIVALENT	41	53.7
Q6PSU3_ARAHY	66933.6	Ara h 1 (7S Globulin)		EQUIVALENT	52	62.8	EQUIVALENT	41	56.9
Q6PSU4_ARAHY	48122.3	Ara h 1 (7S Globulin)	ARA H 2.0201	SUBSET	36	67.5	DIFFERENTIABLE	34	70.6
Q6PSU5_ARAHY	33641.0	Ara h 1 (7S Globulin)		DIFFERENTIABLE	24	55.2	DIFFERENTIABLE	20	51.8
Q6PSU6_ARAHY	34169.2	Ara h 1 (7S Globulin)		DIFFERENTIABLE	27	59.4	DIFFERENTIABLE	20	56.1
ASZ1Q8_9FABA	19202.1	Ara h 2.0101 and 2.0201 (2S Albumin)		SUPERSET	22	81.9	SUBSET	10	63.1
ASZ1Q9_9FABA	20557.7	Ara h 2.0101 and 2.0201 (2S Albumin)	ARA H 2.0101	EQUIVALENT	22	82.6	SUBSET	11	65.7
A8VT50_9FABA	19201.1	Ara h 2.0101 and 2.0201 (2S Albumin)		DIFFERENTIABLE	21	81.2	SUBSET	10	63.1
CONG7p1_ARAHY	20557.7	Ara h 2.0101 and 2.0201 (2S Albumin)		EQUIVALENT	22	82.6	SUBSET	11	65.7
CONG7p2_ARAHY	19202.1	Ara h 2.0101 and 2.0201 (2S Albumin)		SUPERSET	22	81.9	SUBSET	10	63.1
CONG7p3_ARAHY	20238.6	Ara h 2.0101 and 2.0201 (2S Albumin)	ARA H 2.0202	DIFFERENTIABLE	22	83.5	DIFFERENTIABLE	12	67.1
CONG7p4_ARAHY	18883.0	Ara h 2.0101 and 2.0201 (2S Albumin)	ARA H 2.0102	DIFFERENTIABLE	21	82.3	DIFFERENTIABLE	11	64.6
A1DZF0_ARAHY	60680.9	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)	ARA H 3.0101	DIFFERENTIABLE	37	82.4	DIFFERENTIABLE	24	66.7
AHY3_ARAHY	54990.9	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	18	66.3	DIFFERENTIABLE	9	26.7
B5TYU1_ARAHY	60929.0	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	34	81.7	DIFFERENTIABLE	23	66.4
O82580_ARAHY	58542.8	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	25	61.5	DIFFERENTIABLE	15	45.6
Q0GM57_ARAHY	58569.8	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)	ARA H 3.0201	EQUIVALENT	26	65.6	EQUIVALENT	22	67.8
Q516T2_ARAHY	61041.0	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	32	73.6	DIFFERENTIABLE	21	59.9
Q647H3_ARAHY	61836.5	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	33	75.2	DIFFERENTIABLE	21	60.0
Q647H4_ARAHY	61810.3	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		SUPERSET	33	76.7	SUPERSET	22	61.9
Q61WGS_ARAHY	58367.7	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)	ARA H 3.0101	EQUIVALENT	26	65.9	EQUIVALENT	22	68.0
Q9FZ11_ARAHY	60753.9	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		DIFFERENTIABLE	34	76.4	DIFFERENTIABLE	22	60.5
Q9SQH7_ARAHY	61429.9	Ara h 3.0101 and 3.0201 (Ara h 3/4; 11S Globulin)		SUBSET	18	50.4	DIFFERENTIABLE	16	45.7
ASZ1Q5_9FABA	17480.2	Ara h 6 (2S Albumin)	ARA H 6.0101	DIFFERENTIABLE	15	78.6	DIFFERENTIABLE	7	57.9
ASZ1Q6_9FABA	17479.2	Ara h 6 (2S Albumin)		SUPERSET	15	78.6	SUPERSET	7	57.9
ASZ1R0_ARAHY	17479.2	Ara h 6 (2S Albumin)		SUPERSET	15	78.6	SUPERSET	7	57.9
CONG_ARAHY	17479.2	Ara h 6 (2S Albumin)		SUPERSET	15	78.6	SUPERSET	7	57.9
B4XID4_ARAHY	19782.5	Ara h 7.0101 and 7.0201 (2S Albumin)	ARA H 7.0101	DIFFERENTIABLE	6	39.6	DIFFERENTIABLE	4	26.8
Q647G8_ARAHY	18786.2	Ara h 7.0101 and 7.0201 (2S Albumin)		DIFFERENTIABLE	5	30.4	SUBSET	2	17.1
Q9SQH1_ARAHY	18748.1	Ara h 7.0101 and 7.0201 (2S Albumin)		DIFFERENTIABLE	2	13.1			
B1PYZ4_ARAHY	16899.7	Ara h 8.0101 (Pathogenesis-Related Protein)		SUPERSET	3	25.5	EQUIVALENT	3	24.2
B2ZGS2_ARAHY	16909.8	Ara h 8.0101 (Pathogenesis-Related Protein)	ARA H 8.0101	SUBSET	2	17.2	EQUIVALENT	3	24.2
Q0PKR4_ARAHY	16914.7	Ara h 8.0101 (Pathogenesis-Related Protein)		DIFFERENTIABLE	2	15.9			
Q2YH1R1_ARAHY	14376.6	Ara h 8.0101 (Pathogenesis-Related Protein)		SUBSET	2	20.0	EQUIVALENT	3	28.1
Q6VT83_ARAHY	16941.8	Ara h 8.0101 (Pathogenesis-Related Protein)		SUPERSET	3	25.5	EQUIVALENT	3	24.2
B0YIU5_ARAHY	16402.5	Ara h 8.0201 (Pathogenesis-Related Protein)	ARA H 8.0201	DIFFERENTIABLE	13	85.0	DIFFERENTIABLE	7	69.3
B6CEX8_ARAHY	12156.1	Ara h 9.0101 and 9.0201 (Non-Specific Lipid Transfer Protein)	ARA H 9.0101	DIFFERENTIABLE	5	45.7	DIFFERENTIABLE	3	31.9
B6CG41_ARAHY	9504.7	Ara h 9.0101 and 9.0201 (Non-Specific Lipid Transfer Protein)	ARA H 9.0201	DIFFERENTIABLE	5	71.7	DIFFERENTIABLE	2	33.7
Q647G4_ARAHY	15574.2	Ara h 10.0101 and 10.0102 (16 kDa Oleosin)	ARA H 10.0101	DIFFERENTIABLE	4	30.0	SUBSET	2	14.0
Q647G5_ARAHY	17798.4	Ara h 10.0101 and 10.0102 (16 kDa Oleosin)	ARA H 10.0102	DIFFERENTIABLE	7	41.4	DIFFERENTIABLE	4	26.0
Q45W86_ARAHY	14344.8	Ara h 11 (14 kDa Oleosin)	ARA H 11.0101	DIFFERENTIABLE	4	27.0	DIFFERENTIABLE	3	24.8
Q45W87_ARAHY	14298.8	Ara h 11 (14 kDa Oleosin)		DIFFERENTIABLE	4	27.0	DIFFERENTIABLE	3	24.8

SUPPLEMENTAL TABLE 1A

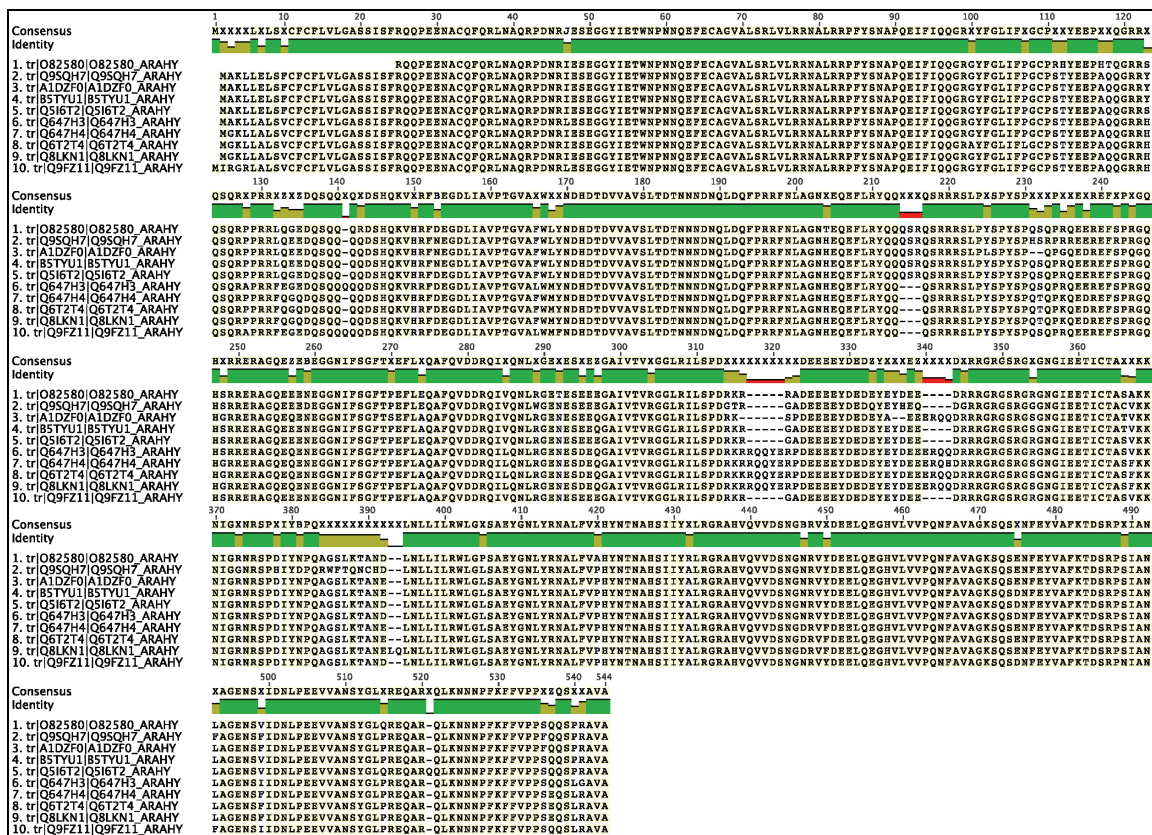
(B) Additional *Arachis hypogaea* Proteins Identified in GELFrEE and Whole Protein Lysate Digestions

Protein Name	Mass	Protein Family	IUIS	GELFrEE FRACTIONATION			LYSATE DIGESTION		
				Parsimony Type	Peptides	% Coverage	Parsimony Type	Peptides	% Coverage
Q4JME7_ARAHY	97819.8	Lipoxygenase		DIFFERENTIABLE	50	90.6	SUPERSET	31	48.6
B4UWB9_ARAHY	26777.7	Lipoxygenase - Fragment		DIFFERENTIABLE	6	51.0			
B4UWC0_ARAHY	20545.5	Lipoxygenase - Fragment		DIFFERENTIABLE	10	80.4	DIFFERENTIABLE	6	41.9
C6ZGE3_9FABA	22348.7	Lipoxygenase - Fragment		DIFFERENTIABLE	2	13.6			
Q2HWT8_ARAHY	92339.5	Phospholipase D		DIFFERENTIABLE	10	15.6			
Q647H1_ARAHY	76343.2	Conarachin (Q647H1_ARAHY)		DISCRETE	21	53.5	DISCRETE	12	25.4
Q20356_ARAHY	51949.2	Ribulose Biphosphate Carboxylase		DIFFERENTIABLE	10	28.6	DIFFERENTIABLE	13	35.7
A7LB59_ARAHY	39857.3	Stereoisom-B (A7LB59_ARAHY)		DISCRETE	4	14.4	DISCRETE	6	19.3
A7LB60_ARAHY	39009.2	Stereoisom-A (A7LB60_ARAHY)		DISCRETE	6	18.3	DISCRETE	6	23.8
A1E2B0_ARAHY	33840.9	11S Seed Storage Globulin (A1E2B0_ARAHY)		DISCRETE	15	44.3	DISCRETE	11	39.3
A1E2B1_ARAHY	31476.0	11S Seed Storage Globulin (A1E2B1_ARAHY)		DISCRETE	5	24.5	DISCRETE	5	26.7
Q43373_ARAHY	29605.0	Galactose-Binding Lectin (Q43373_ARAHY)		DISCRETE	2	6.5			
LECG_ARAHY	29363.8	Galactose-Binding Lectin (LECG_ARAHY)		DIFFERENTIABLE	10	64.1	DIFFERENTIABLE	10	64.1
B4UW79_ARAHY	24398.6	Glutathione Peroxidase		DISCRETE	4	19.4	DISCRETE	3	22.7
B4UWB4_ARAHY	23136.3	Lactoylglutathione Lyase		DISCRETE	3	17.5			
B4UWA1_ARAHY	22383.5	Putative Uncharacterized Protein (B4UWA1_ARAHY)		DISCRETE	5	45.9	DISCRETE	2	11.2
B4UWB2_ARAHY	22079.2	Kunitz Trypsin Inhibitor 4		DISCRETE	2	9.9			
B4UWA3_ARAHY	20364.4	Putative Uncharacterized Protein (B4UWA3_ARAHY)		DISCRETE	5	49.2			
B4UWE2_ARAHY	20302.5	Perchloric Acid Soluble Translation Inhibitor Protein		DISCRETE	4	29.6	DISCRETE	2	15.9
Q06H34_ARAHY	19248.2	Tonoplast Intrinsic Protein Alpha TIP		DISCRETE	5	61.2			
Q06H31_ARAHY	19172.6	Translationally Controlled Tumor-like Protein		DISCRETE	5	42.9	DISCRETE	2	17.9
B4UW70_ARAHY	18775.6	Fiber Annexin		DISCRETE	4	28.6	DISCRETE	9	62.7
Q6J1J8_ARAHY	18493.9	Oleosin 18 kDa		SUPERSET	3	12.5	SUBSUMABLE	2	11.4
Q9AXI0_ARAHY	18502.8	Oleosin 18 kDa		SUBSET	2	6.2	DIFFERENTIABLE	2	10.8
Q9AXI1_ARAHY	18480.9	Oleosin 18 kDa		DIFFERENTIABLE	3	12.5	DIFFERENTIABLE	2	11.4
Q06H21_ARAHY	17868.6	Ubiquitin-Ribosomal Protein S27a		SUPERSET	4	27.7	EQUIVALENT	2	11.6
Q06H40_ARAHY	21432.5	Ubiquitin-Ribosomal Protein S27a		SUPERSET	4	53.1	EQUIVALENT	2	18.8
Q45W78_ARAHY	14892.0	Ubiquitin-Ribosomal Protein S27a		SUPERSET	4	33.6	EQUIVALENT	2	14.1
B4UW81_ARAHY	20386.8	Glutathione S-Transferase 2		EQUIVALENT	3	25.7			
B4UW83_ARAHY	15191.2	Glutathione S-Transferase 2		EQUIVALENT	3	34.6			
Q06H32_ARAHY	17525.2	Thioredoxin Fold		DISCRETE	6	51.9	DIFFERENTIABLE	6	60.5
ATPB_HUMAN	56524.6	ATP Synthase Beta Subunit		DIFFERENTIABLE	3	11.5	EQUIVALENT	2	4.5
D3Y4M7_BRAD	17501.0	ATP Synthase Beta Subunit		EQUIVALENT	2	14.7	EQUIVALENT	2	14.7
D3Y4M9_BRAEL	17501.0	ATP Synthase Beta Subunit		EQUIVALENT	2	14.7	EQUIVALENT	2	14.7
D3Y4N1_RHILE	16956.7	ATP Synthase Beta Subunit		SUBSET	2	15.1	EQUIVALENT	2	15.1
D3Y4N3_BRAD	17818.1	ATP Synthase Beta Subunit		EQUIVALENT	2	14.5	EQUIVALENT	2	14.5
D3Y4N4_BRAEL	17414.0	ATP Synthase Beta Subunit		EQUIVALENT	2	14.8	EQUIVALENT	2	14.8
D3Y4N5_BRAD	17521.1	ATP Synthase Beta Subunit		EQUIVALENT	2	14.7	EQUIVALENT	2	14.7
D3Y4N6_RHIZ	16768.7	ATP Synthase Beta Subunit		SUBSET	2	15.2	EQUIVALENT	2	15.2
D3Y4N7_RHIZ	16841.7	ATP Synthase Beta Subunit		SUBSET	2	15.1	EQUIVALENT	2	15.1
D3Y4N8_RHIZ	16768.7	ATP Synthase Beta Subunit		SUBSET	2	15.2	EQUIVALENT	2	15.2
D3Y4N9_BRAD	17415.9	ATP Synthase Beta Subunit		EQUIVALENT	2	14.8	EQUIVALENT	2	14.8
D3Y4P0_BRAD	17428.0	ATP Synthase Beta Subunit		EQUIVALENT	2	14.8	EQUIVALENT	2	14.8
D3Y4P2_RHIME	16824.7	ATP Synthase Beta Subunit		SUBSET	2	15.1	EQUIVALENT	2	15.1
D3Y4P8_BRAD	17507.0	ATP Synthase Beta Subunit		EQUIVALENT	2	14.7	EQUIVALENT	2	14.7
D5KZX9_RHIZ	17140.8	ATP Synthase Beta Subunit		SUBSET	2	14.7	EQUIVALENT	2	14.8
Q06H19_ARAHY	16908.1	UDP-Glucose Pyrophosphorylase		DISCRETE	8	65.3	DISCRETE	6	42.7
Q6PWX0_ARAHY	16714.8	Calmodulin		SUPERSET	5	42.6			
Q6R2U4_ARAHY	16676.8	Calmodulin		SUPERSET	5	42.6			
Q45W80_ARAHY	16428.6	Nucleoside Diphosphate Kinase		DISCRETE	7	55.7	DISCRETE	4	32.2
Q2PXN9_ARAHY	16159.3	Dessication Protectant Protein LEA 14		DISCRETE	2	19.0			
Q1PCR4_ARAHY	16062.1	Putative IN2-1 Protein		DISCRETE	2	22.3			
Q1HDS7_ARAHY	15302.5	Superoxide Dismutase		DIFFERENTIABLE	3	38.8	EQUIVALENT	2	32.2
Q45W82_ARAHY	15203.4	Superoxide Dismutase		SUBSET	2	32.2	EQUIVALENT	2	32.2
B4UWD5_ARAHY	15118.0	Proteasome Subunit Alpha Type		DIFFERENTIABLE	2	26.3			
D3K177_ARAHY	14207.1	Profilin		DISCRETE	2	16.8			
B3GR01_9FABA	10760.5	Cystatin		EQUIVALENT	2	25.5			
Q647G6_ARAHY	10760.5	Cystatin		EQUIVALENT	2	25.5			
Q4U4M1_ARAHY	10408.1	LEA protein		SUPERSET	7	64.2	SUPERSET	4	52.6
Q850G8_ARAHY	10390.1	LEA protein		DIFFERENTIABLE	7	55.8	SUBSET	3	38.9
IBB1_ARAHY	8426.4	Bowman-Birk Type Proteinase Inhibitor		DIFFERENTIABLE	5	78.6	EQUIVALENT	2	48.6
Q0PKR5_ARAHY	7878.2	Bowman-Birk Type Proteinase Inhibitor		SUBSET	4	56.2	EQUIVALENT	2	53.1
Q2VMU0_ARAHY	12286.6	Bowman-Birk Type Proteinase Inhibitor		DIFFERENTIABLE	4	33.6			
Q7X973_ARAHY	9495.9	Bowman-Birk Type Proteinase Inhibitor		SUBSET	4	45.0	EQUIVALENT	2	42.5
Q0Q0Q8_ARAHY	9148.6	Type 4 Metallothionein (Q0Q0Q8_ARAHY)		DISCRETE	2	24.4			
Q0Q0Q9_ARAHY	9058.6	Type 4 Metallothionein (Q0Q0Q9_ARAHY)		DISCRETE	2	30.5			
Q2PXN4_ARAHY	5801.9	Seed Maturation Protein LEA 4		DISCRETE	2	29.1			

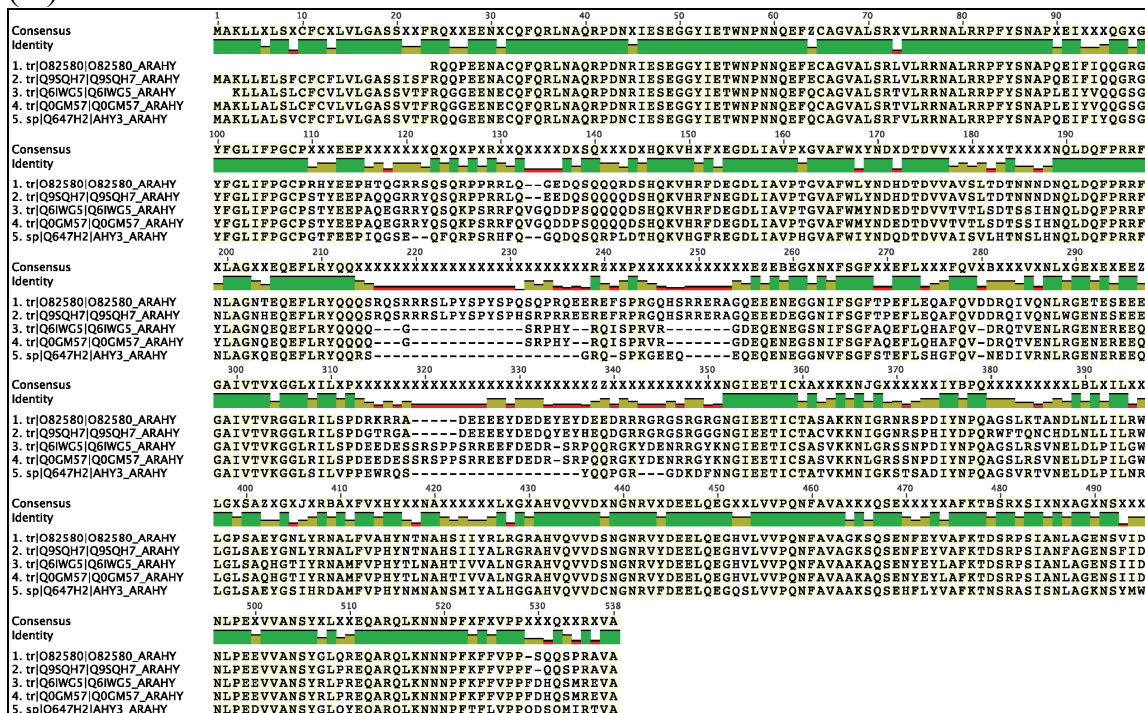
SUPPLEMENTAL TABLE 1B

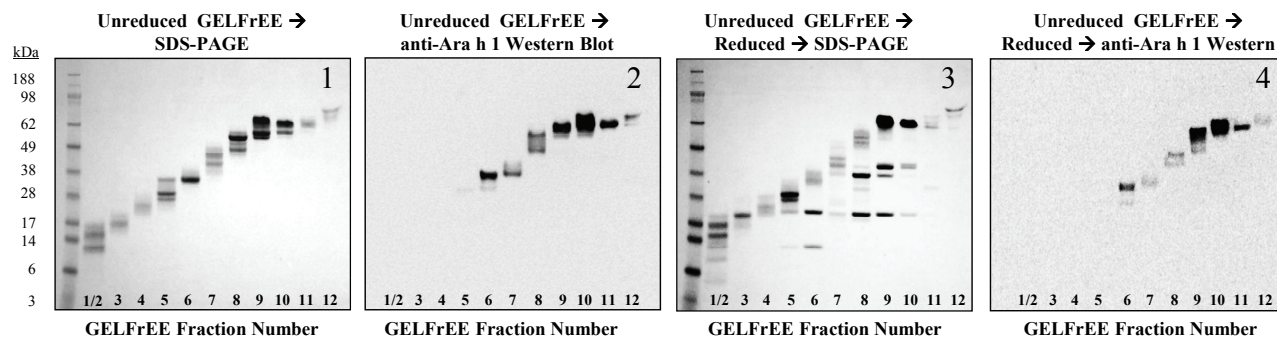


SUPPLEMENTAL FIGURE 1



SUPPLEMENTAL FIGURE 2





SUPPLEMENTAL FIGURE 4



SUPPLEMENTAL FIGURE 5