

Quantitative Proteomic Profiling of Peanut Allergens in Food Ingredients Used for Oral Food Challenges

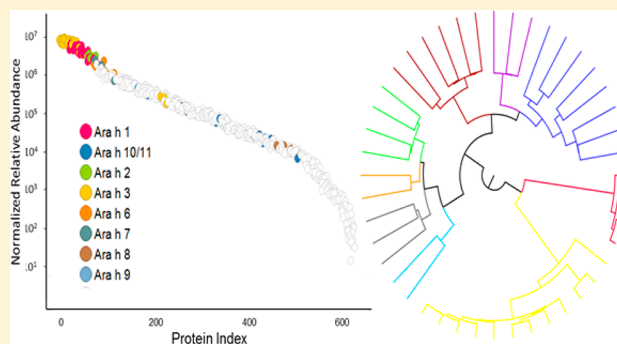
Philip E. Johnson,^{†,§,⊥} Rebekah L. Sayers,^{†,⊥} Lee A. Gethings,^{‡,⊥} Anuradha Balasundaram,[†] Justin T. Marsh,^{†,§} James I. Langridge,[‡] and E. N. Clare Mills^{*,†}

[†]Manchester Institute of Biotechnology, Institute of Inflammation and Repair, Manchester Academic Health Sciences Centre, University of Manchester, Manchester, United Kingdom M17DN

[‡]Waters Corporation, Stamford Avenue, Altrincham Road, Wilmslow, United Kingdom SK9 4AX

S Supporting Information

ABSTRACT: Profiling allergens in complex food ingredients used in oral food challenges and immunotherapy is crucial for regulatory acceptance. Mass spectrometry based analysis employing data-independent acquisition coupled with ion mobility mass spectrometry—mass spectrometry (DIA-IM-MS) was used to investigate the allergen composition of raw peanuts and roasted peanut flour ingredients used in challenge meals. This comprehensive qualitative and quantitative analysis using label-free approaches identified and quantified 123 unique protein accessions. Semiquantitative analysis indicated that allergens Ara h 1 and Ara h 3 were the most abundant proteins and present in approximately equal amounts and were extracted in reduced amounts from roasted peanut flours. The clinically significant allergens Ara h 2 and 6 were less abundant, but relative quantification was unaffected by roasting. Ara h 5 was undetectable in any peanut sample, while the Bet v 1 homologue Ara h 8 and the lipid transfer protein allergen, Ara h 9, were detected in low abundance. The oleosin allergens, Ara h 10 and 11, were moderately abundant in the raw peanuts but were 100-fold less abundant in the defatted roasted peanut flour than the major allergens Ara h 1, 3, 2, and 6. Certain isoforms of the major allergens dominated the profile. The relative quantitation of the major peanut allergens showed little variation between different batches of roasted peanut flour. These data will support future development of targeted approaches for absolute quantification of peanut allergens which can be applied to both food ingredients used in clinical studies and extracts used for skin testing and to identify trace levels of allergens in foods.



IgE-mediated food allergies have been estimated to affect around 2% of adults and up to 4–5% of infants and young children.¹ While allergy to cow's milk and egg in infancy is generally outgrown, allergies to other foods, such as peanut, are usually life-long. Consequently, as there is currently no accepted therapy for food allergy, patients have to practice food avoidance with those considered at risk of suffering a severe reaction being given rescue medication, often in the form of self-injectable adrenaline. Difficulties in managing food avoidance and the fear of severe reactions mean that food allergy impacts adversely on the quality of life of patients and their families.² Receiving an adequate diagnosis is crucial with oral food challenges (open, single, or double blinded) having considerable utility in confirming a diagnosis.^{3,4} While standardized clinical protocols have been proposed and approaches developed to ensure adequate blinding of allergenic food ingredients,^{5–8} standards for allergenic ingredients included in food challenges are still lacking.

One way of demonstrating that a food ingredient has maximal allergenic activity is to show it has a repertoire of allergens representative of those found in commonly consumed

foods associated with triggering allergic reactions. A number of allergens have been identified in peanut flour^{9,10} including the cupin superfamily members which form the major seed storage proteins of peanut known as the 7S and 11S seed storage globulins (Ara h 1 and Ara h 3, respectively). Others belong to the prolamin superfamily and include the 2S albumin fraction (Ara h 2, 6, and 7) and the lipid transfer protein allergen associated with allergy to peach in the Mediterranean basin (Ara h 8/9). Other minor proteins in peanut which are also allergens include those associated with pollen sensitization, such as the Bet v 1 homologue of peanut (Ara h 8) and profilin (Ara h 5), together with the oil-body associated oleosin allergens (Ara h 10 and 11). While direct evidence as to the potency of these individual allergens in allergic patients is lacking, IgE binding studies to specific allergen components, when taking into account geographic differences in sensitization patterns to allergens, have consistently shown that the prolamin

Received: November 25, 2015

Accepted: April 11, 2016

Published: April 11, 2016



superfamily allergens (Ara h 2, 6, 7, and 9) together with the cupin allergens (Ara h 1 and 3) are the most important in eliciting allergic reactions.^{11,12}

Since many food allergens are represented by multiple gene families, tools are required which have the capacity of profiling many protein molecules within a given ingredient. Gel-based methods have been used to profile peanut proteins and allowed mapping of the main allergens.^{13–16} More recently, mass spectrometry (MS) methods have been identified as tools suitable for profiling allergens.¹⁷ Targeted approaches using stable isotope-labeled peptides as internal standards with spectral counting and multiple reaction monitoring approaches have been applied to profiling soybean allergens.¹⁸ However, data-independent analysis (DIA), coupled with ion mobility tandem mass spectrometry (DIA–IM–MS, also known as LC–HDMSE), also offers the means of profiling allergens in complex food ingredients. MSE has been applied to profiling of wheat allergens¹⁹ but has yet to be applied to peanut allergen profiling. This paper describes the profiling of allergens in a lightly roasted mechanically defatted peanut flour ingredient sourced from the United States. This has previously been used as the active ingredient in food challenges^{20,21} and for oral tolerance induction.^{22–24}

MATERIALS AND METHODS

Reagents. All reagents and chemicals were of analytical reagent grade unless otherwise stated; further information on laboratory reagents can be found in the [Supporting Information](#). Raw peanuts were ground and defatted at room temperature with hexane prior to extraction according to Johnson et al.²⁵ Batches of peanut flour “light 12%” were obtained from Byrd Mill (Byrd Mill Co., Ashland, VA 23005). The flour was manufactured by the Golden Peanut Company, LLC, Alpharetta, GA 30022 and is a lightly roasted mechanically defatted peanut flour with a minimum 50% (w/w) protein. Batch 1 was an archival samples used for clinical studies in the EuroPrevall project,^{21,26} and the remaining batches having been sourced over a 4 year period and stored under ambient conditions in the dark. Batch 3 was used to prepare a quality control material for allergen analysis (Holcombe, Walker, Gethings, Johnson, and Mills, unpublished work). The total protein nitrogen of raw and roasted peanut flours was determined by Kjeldahl analysis according to the AOCS Ab 4-91 method²⁷ by Alliance Technical (Needham, Suffolk, U.K.) using a factor of 5.46 for conversion of N to protein.²⁸

Sample Preparation. Peanut samples (20 mg) were extracted in triplicate by heating at 60 °C for 15 min with continuous sonication in a sonicating water bath using 50 volumes (w:v) in either Tris/DTT (50 mM Tris-HCl pH 8.8, containing 50 mM DTT and 0.04% (w/v) Rapigest) or a chaotropic-zwitterionic extraction buffer (7 M urea, 2 M thiourea, 2% (w/v) CHAPS; 1% (w/v) DTT). Extracts were clarified by centrifugation at 10 000g for 10 min, supernatants removed and frozen at –20 °C until required. The protein content of extracts was determined with a 2-D Quant protein assay kit (GE Healthcare Life Sciences, Buckinghamshire, U.K.) according to manufacturer's instructions using bovine serum albumin as a standard. Samples were analyzed by SDS-PAGE and immunoblotting ([Supporting Information](#)).

Mass Spectrometry. Tris/DTT extracts were diluted to 1 mg/mL protein with 50 mM ammonium bicarbonate and reduced by addition of DTT to 10 mM. Samples were

incubated at 80 °C for 10 min prior to alkylation by the addition of iodoacetamide to 20 mM. After incubating for 20 min in the dark at ambient temperature, trypsin was added to 3 µg/mL and samples digested for 2 h at 37 °C followed by the further addition of another 3 µg/mL trypsin and digestion overnight at 31 °C. Following digestion, formic acid was added to a final concentration of 0.1% (v/v) which was sufficient to hydrolyze the low level of Rapigest used, and samples were clarified by centrifugation at 10 000g for 10 min. Supernatants were frozen at –20 °C prior to analysis by mass spectrometry (MS). One-dimensional (1D) nanoscale liquid chromatography separation of tryptic peptides was performed with a nano-Acquity Ultra Performance liquid chromatography (UPLC) system (Waters Corporation, Milford, MA), equipped with a Symmetry C18 5 µm, 2 cm × 180 µm precolumn and HSS T3 C18 1.8 µm, 15 cm × 75 µm analytical reversed phase column (Waters Corporation, Milford, MA). Prior to analysis, a standard digest of glycogen phosphorylase B (P00489) was added to each sample prior to LC/MS analysis, to give 5 fmol/µL. For detailed chromatographic and mass spectrometry conditions, please refer to the [Supporting Information](#).

Analysis of LC–MS Data. Progenesis Q1 for proteomics informatics (originally TransOmics informatics Nonlinear Dynamics, Newcastle, U.K.) was used to process all mass spectral data acquired. Protein identifications were obtained by searching a sequence database of the peanut (*Arachis hypogaea*, taxon identifier 3818) proteome as present in UniProt (<http://www.uniprot.org/>,²⁹) on 8/7/2015; 1131 sequences in total. Please refer to [Supporting Information](#) for further details on database preparation and data analysis.

Statistical Comparisons. Linear regressions and statistical descriptors including a spearman correlation for comparison of roasted flour batches were generated using GraphPad Prism (Version 6.04) using recommended settings.

RESULTS AND DISCUSSION

Initially the effectiveness of two different extraction procedures was compared using three batches of raw peanuts and two different batches of commercial roasted peanut flour. Extraction efficiency was calculated relative to the total protein on the basis of Kjeldahl nitrogen analysis ([Table S-1](#)). Additionally, the presence of major allergenic proteins in different extracts was confirmed with SDS-PAGE and immunoblotting using allergen specific antibody preparations ([Figure S-1](#)). On the basis of these experiments, an extraction with Tris/DTT and acid-labile detergent was judged to give a representative protein extract while minimizing required sample cleanup and was therefore selected for further MS analysis.

Proteins were identified using a combination of assumed peptide masses from low collision energy *m/z* scans (equivalent to parent ion mass in a data-dependent acquisition (DDA) mode combined with fragment ions masses from elevated collision energies (equivalent to MS2 in a DDA workflow))^{30,31} ([Scheme S-1](#)). A key difference between DIA–IM–MS and data-dependent acquisition is the lack of serial parent ion selection for MS2 fragmentation, a feature which has been shown to provide more robust quantitation of the MS signal.³² Compared with more traditional DDA acquisition schemas, a DIA label-free approach was adopted to limit issues regarding peptide sampling reproducibility and identification rates, which are due to the stochastic nature and sampling frequency of DDA workflows.³³ Increased peak capacity and selectivity are

achieved by implementing DIA in combination with ion mobility.

Prior to performing the analysis of LC–MS data, a curated set of relevant peanut protein sequences was compiled. This list of peanut allergens was compiled by undertaking homology searches using the UniProt protein sequence database for peanut (Taxid 3818) with exemplar allergen sequences from the IUIS nomenclature committee database (Tables S-2 and S-3). A phylogenetic comparison of these sequences (Figure 1;

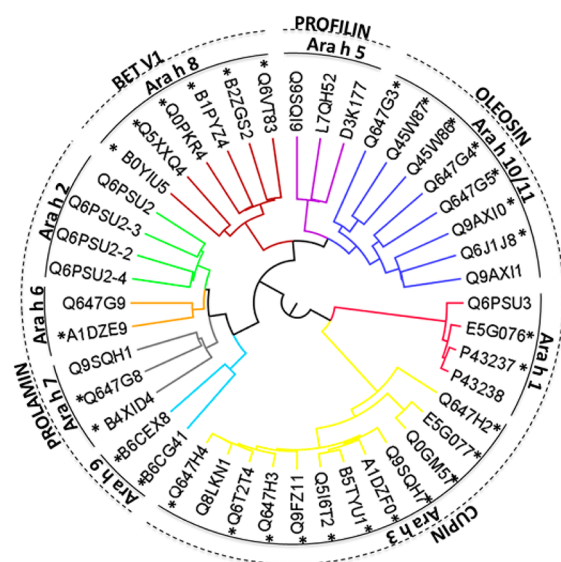


Figure 1. Phylogenetic comparison of peanut allergen sequences and their identification by DIA–IM–MS. Sequences of collated allergens (Table S-3) were aligned (Figures S-2–5), for major allergens Ara h 1, 3, 2, 5, 6, 7, 8, 9, 10, and 11 and annotated with UniProt accession numbers. IUIS allergen names are shown above the solid circular line, and protein family names are shown above the dotted line. A bold line on the dendrogram indicates the allergen sequence was identified by LC–MS based on all (unique and shared) peptides; an asterisk denotes detection based on peptides unique to the indicated allergen sequence (Table S-3).

alignments shown for major allergens Ara h 1, 3, 2, and 6 in Figures S-2–5) shows a clear homology-based division into the allergen superfamilies, which is dominated by the cupin (Ara h 1 and 3) and prolamin (Ara h 2, 6, 7 and 9) superfamilies with three other families, the profilin (Ara h 5), Bet v 1 (Ara h 8) and oleosin (Ara h 10 and 11) families also represented. The IUIS allergen nomenclature subcommittee (www.allergen.org) recognizes Ara h 12 and Ara h 13 as peanut allergens, but these are listed only as expressed sequence tag (EST) sequences. UniProt does not contain protein sequences with significant homology to these EST sequences; therefore, Ara h 12 and Ara h 13 were not considered in this study. Each family is represented by multiple protein sequences reflecting their multigene character, which needs to be considered in MS-based detection and quantitation of peanut allergens. For example, 7 and 14 nonredundant sequences were identified for Ara h 1 and Ara h 3, respectively, of which 3 from Ara h 1 and two from Ara h 3 were lacking signal sequences and hence will not be expressed and deposited in the protein bodies of the seeds (Figures S2 and S3). For this reason these sequences were excluded from the final sequence sets used for analysis. The average pairwise identities of sequences within these allergen families ranged from 73.7% (Ara h 10 and 11) to 95.2% (Ara h

1). This set of 45 sequences was subsequently used for MS data analysis (Table S-3).

DIA–IM–MS-based label-free quantitation identified around 123 unique protein accessions from tryptic digests of peanut extracts, including the major protein families associated with peanut allergy which were identified with sequence coverage of 60% or greater (Figure 2; Tables S-3 and S-4). A number of the proteins have been identified previously in peanut seeds through 2D-PAGE proteomic profiling, including lipoxygenases, a variety of metabolic proteins together with many thought to play a role in plant protection. The N-terminal precursor region of Ara h 1 was identified as being present in all samples with good sequence coverage (Figure S-2B). Although it has never been purified from peanut, the corresponding fragment of the walnut allergen Jug r 1 has been identified as an allergen.³⁴ Normalized relative abundance for all proteins identified in raw and roasted peanut samples spanned 2–3 orders of magnitude and provided evidence for the existence of all the allergen sequences apart from Ara h 5, with 23 isoforms inferred from unique peptide matches (Figure 2; Table S-3). Sequence coverage was greater for proteins identified in raw peanuts, although the majority of additional peptides identified had missed cleavages, probably because the denaturing effect of roasting enhanced the efficiency of trypsin digestion (Table S-3). Prolyl hydroxylation was a widely found post-translational modification (Figures S-2–5) in all allergens and included those previously observed in Ara h 2³⁵ and have been shown to play an important role in IgE binding in Ara h 6.³⁶

Semiquantification of allergens using spectral counting was also undertaken on an isoform-by-isoform basis (Table S-3; Figure 3A). Consistent with the normalized relative abundance analysis, Ara h 1 and Ara h 3 were found to be the most abundant proteins. This is consistent with their being the major seed storage proteins of peanut, constituting around 50% of the seed protein. The prolamin superfamily allergens Ara h 2, 6, and 7 were less abundant in the raw peanuts, consistent with the observations of others,³⁷ but were unaffected by roasting, with Ara h 9, the lipid transfer protein (LTP) allergen of peanut, being a minor component. This is likely due to the fact the skins were removed from both the raw peanut seeds and the peanuts used to prepare the roasted peanut flour, reflecting the fact that peanuts are generally consumed after blanching. The oleosin allergens Ara h 10 and 11 are typically differentiated by difference in migration by SDS-PAGE (M_r 16 and 18 kDa, respectively) but share high sequence homology. The major observed oleosin allergens Ara h 10 (Q647G5) and Ara h 11 (Q9AX10 and Q6JJJ8) were originally identified from oil body fractions of peanut seeds.³⁸ The abundance of Ara h 11 isoforms was reduced in the roasted flour compared to the whole raw peanuts and reflects the fact that it is an oil body associated protein, and the oil bodies have largely been lost during the mechanical defatting process used to prepare the flour used in this study. However, this reduction was much less marked for Ara h 10.

All isoforms of the pollen-related Bet v 1 allergen Ara h 8 could be identified in the raw peanut samples, although two isoforms (Q6VT83 and B2ZGS2) were represented by a single, unique peptide. Their relative abundance was reduced more than 10-fold in the roasted flour with markedly reduced sequence coverage (Table S-3). No evidence of another pollen-related profilin allergen of peanut, Ara h 5, was found, an observation consistent with its being expressed at very low

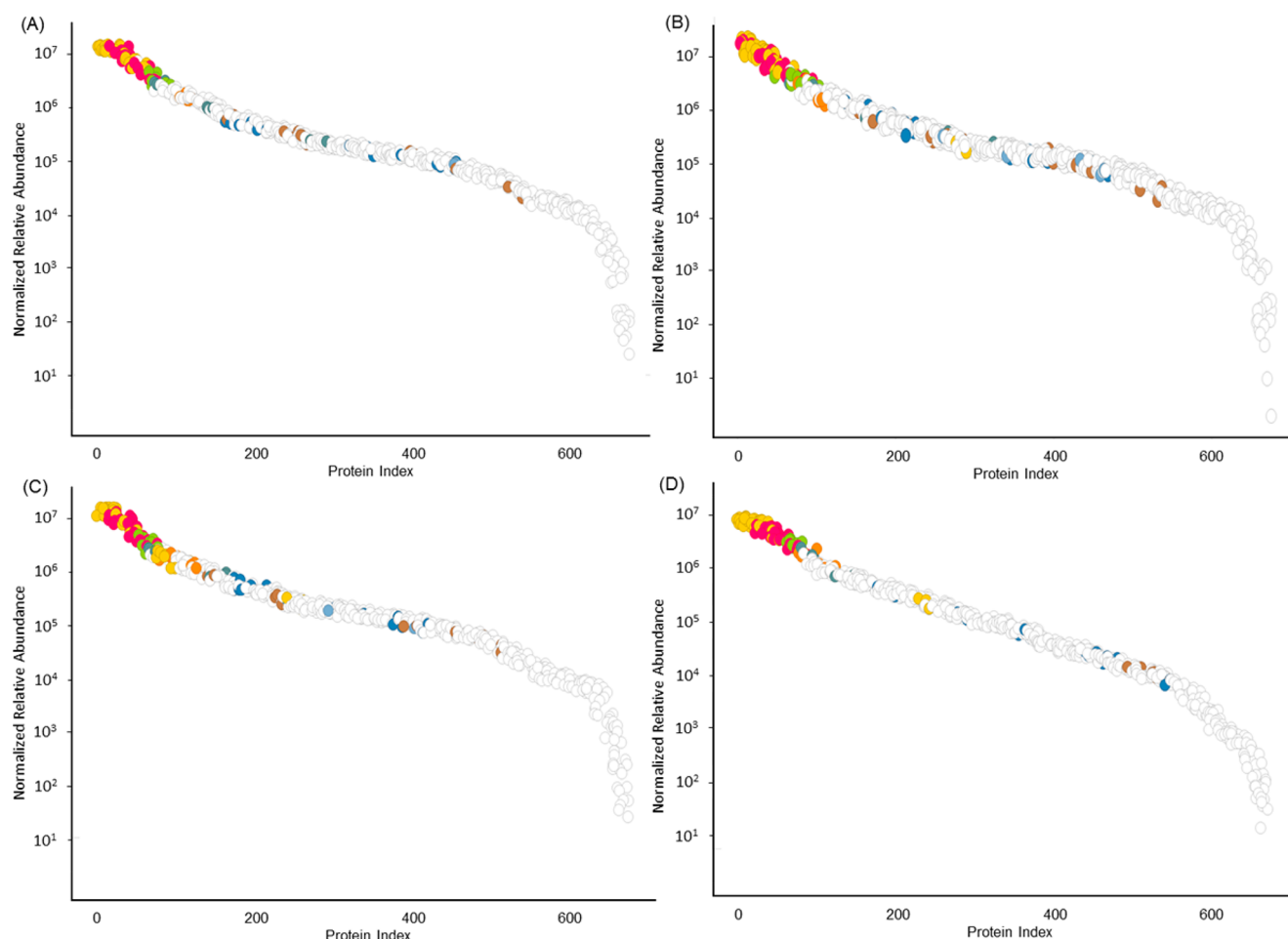


Figure 2. Normalized quantity distribution of identified peanut proteins (including allergens) as determined by DIA–IM–MS: (A) raw sample 1; (B) raw sample 2; (C) raw sample 3; (D) roasted peanut flour batches 1 and 2. Protein index quantitation curves are based on all identifications over all technical replicates ($n = 3$ extracts per sample and each extract analyzed in triplicate). Clear circles represent nonallergenic protein sequences (Table S-4). Filled circles represent peanut allergens (Table S-3) as follows: Ara h 1, red; Ara h 3, yellow; Ara h 2, green; Ara h 6, orange; Ara h 7, gray; Ara h 10/11, dark blue; Ara h 8, brown, and Ara h 9, light blue.

levels in seeds.³⁹ There is evidence from the model plant *Arabidopsis thaliana* that certain forms of profilin are only expressed in pollen, with different isoforms being expressed in vegetative tissues.⁴⁰ Their role in triggering food allergies is unclear and may reflect their low abundance in plant tissues consumed as food, such as seeds.

Attempts to undertake semiquantification of allergen isoforms was not possible because in some cases very few unique peptides were identified or unique tryptic peptides could not be generated from a sequence. This is particularly evident with Ara h 2, where the high degree of homology between isoforms meant that no isoform specific peptides were observed.

Semiquantitation of the major allergens involved in clinical allergy to peanut (Ara h 1, Ara h 3, Ara h 2/6) using all the peptides identified belonging to the different allergen types was also undertaken (Figure 3B.) The cupin allergens Ara h 1 and Ara h 3 were more abundant in the raw and blanched peanut flours than in the roasted flour samples ($P < 0.05$), either because extraction of these allergens was reduced after thermal processing because of aggregation or because of processing-induced modifications. The 2S albumin allergens were present in the same amount in all flour samples ($P > 0.05$).

Using this approach, the allergen content of two different batches of roasted peanut flour was compared (Figure 4).

These data demonstrate the comparability of the different flour batches and allow the generation of statistical comparators describing the relationship between the allergen composition of a given peanut preparation with a comparator. An example of this is shown in Table 1, in which such regression comparators are shown for 6 roasted peanut flour batches, 3 raw peanut flours against roasted peanut flour batch 3.

CONCLUSIONS

An effective extraction procedure is an important first step of any MS-based analysis, particularly of foodstuffs where complex matrices and processing conditions may hinder effective recovery of proteins.⁴¹ We developed a rapid extraction method for peanut proteins from roasted peanut flour and applied it to LC–MS profiling of peanut allergens. While less effective than the chaotropic-zwitterionic detergent buffer, it reproducibly extracted a representative repertoire of proteins and was therefore considered a suitable starting place for the analysis. These data further illustrate the difficulties in developing complete extraction procedures for processed foods which are compatible with MS-based detection methods without the use of relatively complex cleanup procedures.⁴²

The protein-level evidence for presence of particular allergen isoforms has, until now, been largely lacking and hence not

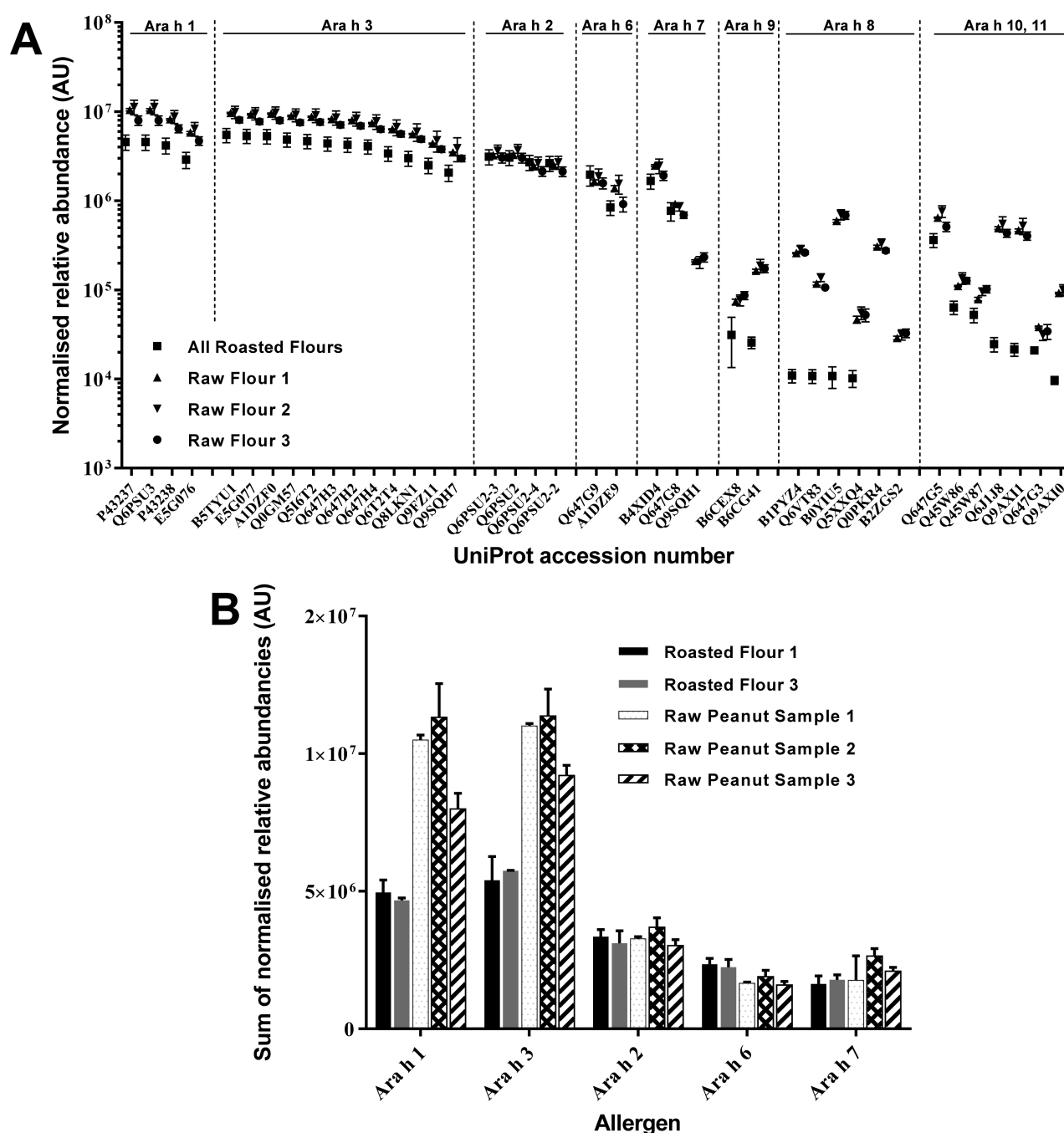


Figure 3. Label-free quantitation of selected allergen sequences undertaken using spectral counting. (A) Major peanut allergens from raw peanuts (raw sample 1, triangle; raw sample 2, inverted triangle), raw sample 3 (circle) and 7 roasted peanut flour batches (filled square) using all peptides (common and unique) to allergen isoforms. Data are expressed as the mean ($\pm$ standard error) for three extracts of each flour. (B) Clinically important allergens in three different samples of raw peanuts and two batches of roasted peanut flour. Data are expressed as the mean ($\pm$ standard error) for ($n = 3$ extracts per sample and each extract analyzed in triplicate). Data in (B) were calculated by summing the total abundance of peptides corresponding to the allergen isoform with the greatest sequence together with the unique peptides from other allergen isoforms. Ara h 1, P43237; Ara h 3, A1DZF0; Ara h 2, Q6PSU2; Ara h 6, Q647G9; Ara h 7, BUX1D4; Ara h 9, B6CEX8.

been available to guide the choice of allergen sequences used for molecular studies or inclusion of allergen components in diagnostic arrays. Both the abundant isoforms of Ara h 2 (Q6PSU2) and Ara h 6 (Q647G9) have been used to define the clinical reactivity of these allergens.⁴³ One allergen, Ara h 5, was not detected in any peanut sample. However, further studies using other extraction procedures and heavy labeled peptides should be undertaken to confirm this observation before considering whether it plays a role in clinical allergy to

peanut seeds. During this study it became apparent that label-free quantitation of proteins of multigene families (i.e., those which share common peptides used for quantitation) may not be possible or provide only limited information where isoforms are highly conserved and only a limited number of unique mass events can be identified. It is also limited by the lack of sequenced genomes for edible crop plants since it may be possible that important highly expressed isoforms have yet to be sequenced.

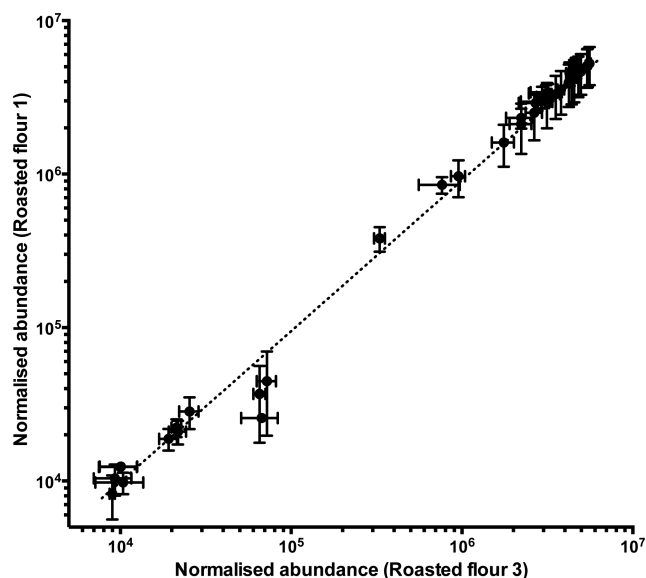


Figure 4. Comparison of normalized relative abundances of allergenic proteins in roasted peanut flour batch 3 using roasted peanut flour batch 1 as a reference.

Table 1. Linear Regression Analysis of x - y Scatter Plots Generated Using Allergen Quantitation in Three Samples of Raw Peanuts and Seven Batches of Roasted Peanut Flour, versus a Standard Comparator (Roasted Peanut Flour Batch 3) Generated Using Either Peptides Unique to Allergen Sequences or Using All Observed Peptides^a

flour batch	unique peptides only		unique and common peptides	
	r^2	sy.x	r^2	sy.x
roasted 1	0.9919	11532	0.9896	201431
roasted 2	0.9863	17245	0.9987	96091
roasted 3	1.000	0.0	1	0
roasted 4	0.9932	10925	0.9880	230616
roasted 5	0.9951	6100	0.9966	98402
roasted 6	0.9934	8503	0.9894	175712
roasted 7	0.9971	5688	0.9933	134063
raw 1	0.6594	201826	0.9144	1.130e+006
raw 2	0.6867	211517	0.9114	1.208e+006
raw 3 (blanched)	0.6092	182560	0.9363	790266

^aThe r^2 (linear correlation coefficient) and sy.x (standard deviation of residuals) for each regression are shown.

DIA-IM-MS has proven to be a technique suitable for profiling allergens in food ingredients used for preparation of oral food challenge meals and for oral immunotherapy. The ability to rapidly generate single statistical descriptors for the composition of allergens in a sample compared to standardized comparator(s) allows for the setting of acceptance/rejection criteria based on standardized analysis of the clinically relevant molecules in preparations for diagnosis and testing. We believe that such knowledge-based comparisons are essential in moving allergen preparations for clinical use onto a sound biopharmaceutical footing. A further advantage of using untargeted methods for analysis of batch-to-batch variation is that comparisons can be updated to include newly discovered allergens as sequences become available simply by reanalyzing existing data.

These same approaches could also be applied to the profiling of extracts used in oral immunotherapy (OIT): skin prick testing reagents and to identify trace contaminations. These data also provide a firm foundation for the design and development of rapid targeted approaches to allergen profiling such as multiple reaction monitoring (LC-MRM) experiments that potentially offer a means of absolute quantitation, although this will require access to reference materials, which are currently lacking for allergen analysis in general.

■ ASSOCIATED CONTENT

§ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.analchem.5b04466](https://doi.org/10.1021/acs.analchem.5b04466).

Detailed description of mass spectrometry methodology, data on extraction efficiency, identified peanut proteins, and alignments of curated allergen sequences and peptides identified in the study; quantitative proteomic profiling of peanut allergens in food ingredients used for oral food challenges (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: clare.mills@manchester.ac.uk. Fax no: 44-(0) 1613065199.

Present Address

[§]The Food Allergy Research and Resource Programme, The University of Nebraska, Lincoln.

Author Contributions

[†]These authors contributed equally to data generation and analysis.

Notes

The authors declare the following competing financial interest(s): Mills is a founder shareholder of a spin-out company, Reacta Biotech Ltd., which is developing oral food challenges for diagnosis of food allergy. Johnson and Marsh are co-inventors of the IP for oral food challenges exploited by Reacta Biotech. Langridge and Gethings and are employees of Waters Corporation.

■ ACKNOWLEDGMENTS

This work was partly funded through a Biotechnology and Biological Sciences Research Council-Technology Strategy Board grant number 101130 (Allergen analysis-developing integrated approaches) and through the Biotechnology and Biological Sciences Research Council through a Cooperative Award in Science and Engineering with Campden BRI to R.L.S.

■ REFERENCES

- (1) Nwaru, B. I.; Hickstein, L.; Panesar, S. S.; Muraro, A.; Werfel, T.; Cardona, V.; Dubois, A. E.; Halken, S.; Hoffmann-Sommergruber, K.; Poulsen, L. K.; Roberts, G.; Van Ree, R.; Vlieg-Boerstra, B. J.; Sheikh, A.; Group, E. F. A. A. G.. *Allergy* **2014**, *69*, 62–75.
- (2) van der Velde, J.; Dubois, A. J.; Flokstra-de Blok, B. J. *Curr. Allergy Asthma Rep.* **2013**, *13*, 651–661.
- (3) Sampson, H. A.; Gerth van Wijk, R.; Bindslev-Jensen, C.; Sicherer, S.; Teuber, S. S.; Burks, A. W.; Dubois, A. E. J.; Beyer, K.; Eigenmann, P. A.; Spergel, J. M.; Werfel, T.; Chinchilli, V. M. *J. Allergy Clin. Immunol.* **2012**, *130*, 1260–1274.
- (4) Muraro, A.; Werfel, T.; Hoffmann-Sommergruber, K.; Roberts, G.; Beyer, K.; Bindslev-Jensen, C.; Cardona, V.; Dubois, A.; duToit,

- G.; Eigenmann, P.; Fernandez Rivas, M.; Halken, S.; Hickstein, L.; Host, A.; Knol, E.; Lack, G.; Marchisotto, M. J.; Niggemann, B.; Nwaru, B. I.; Papadopoulos, N. G.; Poulsen, L. K.; Santos, A. F.; Skypala, I.; Schoepfer, A.; Van Ree, R.; Venter, C.; Worm, M.; Vlieg-Boerstra, B.; Panesar, S.; de Silva, D.; Soares-Weiser, K.; Sheikh, A.; Ballmer-Weber, B. K.; Nilsson, C.; de Jong, N. W.; Akdis, C. A. *Allergy* **2014**, *69*, 1008–1025.
- (5) Vassilopoulou, E.; Douladiris, N.; Sakellariou, A.; Cortes, S. V.; Sinaniotis, A.; Rivas, M. F.; Papadopoulos, N. G. *J. Hum Nutr Diet* **2010**, *23*, 544–549.
- (6) Vlieg-Boerstra, B. J.; Bijleveld, C. M. A.; van der Heide, S.; Beusekamp, B. J.; Wolt-Plompen, S. A. A.; Kukler, J.; Brinkman, J.; Duiverman, E. J.; Dubois, A. E. J. *J. Allergy Clin. Immunol.* **2004**, *113*, 341–346.
- (7) Ronteltap, A.; van Schaik, J.; Wensing, M.; Rynja, F. J.; Knulst, A. C.; de Vries, J. H. *Allergy* **2004**, *59*, 457–460.
- (8) Cochrane, S. A.; Salt, L. J.; Wantling, E.; Rogers, A.; Coutts, J.; Ballmer-Weber, B. K.; Fritsche, P.; Fernandez-Rivas, M.; Reig, I.; Knulst, A.; Le, T. M.; Asero, R.; Beyer, K.; Golding, M.; Crevel, R.; Clare Mills, E. N.; Mackie, A. R. *Allergy* **2012**, *67*, 107–113.
- (9) 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. *Mol. Nutr. Food Res.* **2008**, *52* (Suppl 2), S272–285.
- (10) White, B. L.; Gokce, E.; Nepomuceno, A. I.; Muddiman, D. C.; Sanders, T. H.; Davis, J. P. *J. Agric. Food Chem.* **2013**, *61*, 3957.
- (11) Nicolaou, N.; Murray, C.; Belgrave, D.; Poorafshar, M.; Simpson, A.; Custovic, A. *J. Allergy Clin. Immunol.* **2011**, *127*, 684–685.
- (12) Krause, S.; Reese, G.; Randow, S.; Zennaro, D.; Quarantino, D.; Palazzo, P.; Ciardiello, M. A.; Petersen, A.; Becker, W. M.; Mari, A. *J. Allergy Clin. Immunol.* **2009**, *124*, 771–778.
- (13) Chassaigne, H.; Tregoeat, V.; Norgaard, J. V.; Maleki, S. J.; van Hengel, A. J. *J. Proteomics* **2009**, *72*, 511–526.
- (14) Kottapalli, K. R.; Payton, P.; Rakwal, R.; Agrawal, G. K.; Shibato, J.; Burow, M.; Puppala, N. *Plant Sci.* **2008**, *175*, 321–329.
- (15) Schmidt, H.; Gelhaus, C.; Latendorf, T.; Nebendahl, M.; Petersen, A.; Krause, S.; Leippe, M.; Becker, W. M.; Janssen, O. *Proteomics* **2009**, *9*, 3507–3521.
- (16) Hebling, C. M.; Ross, M. M.; Callahan, J. H.; McFarland, M. A. *J. Proteome Res.* **2012**, *11*, 5384–5395.
- (17) Fernandez, A.; Mills, E. N.; Lovik, M.; Spoek, A.; Germini, A.; Mikalsen, A.; Wal, J. M. *Food Chem. Toxicol.* **2013**, *62*, 1–6.
- (18) Houston, N. L.; Lee, D. G.; Stevenson, S. E.; Ladics, G. S.; Bannon, G. A.; McClain, S.; Privalle, L.; Stagg, N.; Herouet-Guicheney, C.; MacIntosh, S. C.; Thelen, J. J. *J. Proteome Res.* **2011**, *10*, 763–773.
- (19) Uvackova, L.; Skultety, L.; Bekesova, S.; McClain, S.; Hajdud, M. *J. Proteome Res.* **2013**, *12*, 4862–4869.
- (20) Nicolaou, N.; Poorafshar, M.; Murray, C.; Simpson, A.; Winell, H.; Kerry, G.; Harlin, A.; Woodcock, A.; Ahlstedt, S.; Custovic, A. *J. Allergy Clin. Immunol.* **2010**, *125*, 191–197.
- (21) Mackie, A.; Knulst, A.; Le, T. M.; Bures, P.; Salt, L.; Mills, E. N.; Malcolm, P.; Andreou, A.; Ballmer-Weber, B. K. *Mol. Nutr. Food Res.* **2012**, *56*, 1708–1714.
- (22) Varshney, P.; Jones, S. M.; Scurlock, A. M.; Perry, T. T.; Kemper, A.; Steele, P.; Hiegel, A.; Kamilaris, J.; Carlisle, S.; Yue, X.; Kulis, M.; Pons, L.; Vickery, B.; Burks, A. W. *J. Allergy Clin. Immunol.* **2011**, *127*, 654–660.
- (23) Clark, A. T.; Islam, S.; King, Y.; Deighton, J.; Anagnostou, K.; Ewan, P. W. *Allergy* **2009**, *64*, 1218–1220.
- (24) Anagnostou, K.; Islam, S.; King, Y.; Foley, L.; Pasea, L.; Bond, S.; Palmer, C.; Deighton, J.; Ewan, P.; Clark, A. *Lancet* **2014**, *383*, 1297–1304.
- (25) Johnson, P. E.; Van der Plancken, I.; Balasa, A.; Husband, F. A.; Grauwet, T.; Hendrickx, M.; Knorr, D.; Mills, E. N.; Mackie, A. R. *Mol. Nutr. Food Res.* **2010**, *54*, 1701–1710.
- (26) Ballmer-Weber, B. K.; Lidholm, J.; Fernandez-Rivas, M.; Seneviratne, S.; Hanschmann, K. M.; Vogel, L.; Bures, P.; Fritsche, P.; Summers, C.; Knulst, A. C.; Le, T. M.; Reig, I.; Papadopoulos, N. G.; Sinaniotis, A.; Belohlavkova, S.; Popov, T.; Kralimarkova, T.; de Blay, F.; Purohit, A.; Clausen, M.; Jedrzejczak-Czechowicz, M.; Kowalski, M. L.; Asero, R.; Dubakienė, R.; Barreales, L.; Clare Mills, E. N.; van Ree, R.; Vieths, S. *Allergy* **2015**, *70*, 391–407.
- (27) In AOCS Official Method Ba Ai 4-91; Firestone, D. E., Ed.; AOCS Press: Champaign, IL, 1997.
- (28) Misra, J. B. *Peanut Sci.* **2001**, *28*, 48–51.
- (29) Uniprat Consortium. *Nucleic Acids Res.* **2014**, *42*, D191–D198.
- (30) Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G. Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P.; Geromanos, S. *Anal. Chem.* **2005**, *77*, 2187–2200.
- (31) Bateman, R. H.; Carruthers, R.; Hoyes, J. B.; Jones, C.; Langridge, J. I.; Millar, A.; Vissers, J. P. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 792–803.
- (32) Cheng, F. Y.; Blackburn, K.; Lin, Y. M.; Goshe, M. B.; Williamson, J. D. *J. Proteome Res.* **2009**, *8*, 82–93.
- (33) Michalski, A.; Cox, J.; Mann, M. *J. Proteome Res.* **2011**, *10*, 1785–1793.
- (34) Downs, M. L.; Semic-Jusufagic, A.; Simpson, A.; Bartra, J.; Fernandez-Rivas, M.; Rigby, N. M.; Taylor, S. L.; Baumert, J. L.; Mills, E. N. *J. Agric. Food Chem.* **2014**, *62*, 11767–11775.
- (35) Li, J.; Shefcheck, K.; Callahan, J.; Fenselau, C. *Protein Sci.* **2010**, *19*, 174–182.
- (36) Bernard, H.; Guillon, B.; Drumare, M.-F.; Paty, E.; Dreskin, S. C.; Wal, J.-M.; Adel-Patient, K.; Hazebrout, S. *J. Allergy Clin. Immunol.* **2015**, *135*, 1267.
- (37) Koppelman, S. J.; Vlooswijk, R. A.; Knippels, L. M.; Hessing, M.; Knol, E. F.; van Reijssen, F. C.; Bruijnzeel-Koomen, C. A. *Allergy* **2001**, *56*, 132–137.
- (38) Pons, L.; Chery, C.; Romano, A.; Namour, F.; Artesani, M. C.; Gueant, J. L. *Allergy* **2002**, *57* (Suppl72), 88–93.
- (39) Kleber-Janke, T.; Cramer, R.; Scheurer, S.; Vieths, S.; Becker, W. M. *J. Chromatogr., Biomed. Appl.* **2001**, *756*, 295–305.
- (40) Kandasamy, M. K.; McKinney, E. C.; Meagher, R. B. *Cell Motil. Cytoskeleton* **2002**, *52*, 22–32.
- (41) Johnson, P. E.; Marsh, J. T.; Mills, E. N. C. *Quantitative Proteomics* **2014**, 329.
- (42) Shefcheck, K. J.; Callahan, J. H.; Musser, S. M. *J. Agric. Food Chem.* **2006**, *54*, 7953–7959.
- (43) Lehmann, K.; Schweimer, K.; Reese, G.; Randow, S.; Suhr, M.; Becker, W. M.; Vieths, S.; Rosch, P. *Biochem. J.* **2006**, *395*, 463–472.

Quantitative proteomic profiling of peanut allergens in food ingredients used for oral food challenges

Philip E. Johnson^{1†‡}, Rebekah L. Sayers^{1‡}, Lee A. Gethings^{2‡}, Anuradha Balasundaram¹, Justin T. Marsh^{1†}, James I. Langridge² and E.N.Claré Mills^{1*}.

¹Manchester Institute of Biotechnology, Institute of Inflammation and Repair, Manchester Academic Health Sciences Centre, University of Manchester, UK, M17DN.

²Waters Corporation, Stamford Avenue, Altrincham Road, Wilmslow, UK SK9 4AX.

Supporting Information

Contents

Supplementary Materials and Methods	3
Figure S-1. SDS-PAGE analysis and immunoblots of raw and roasted peanut flour extracts prepared using two different buffers.....	5
Table S-1. Effect of roasting on peanut protein extraction efficiency using two different buffers.	6
Scheme S-1. Workflow for the analysis of MS data using an initial screen of all peanut protein sequences (UniProt) and further analysis of selected allergen sequences.....	7
Table S-2. Additional information on the detection of the most abundant allergen isoforms in this study.	8
Figure S-2 A. Alignment of Ara h 1 curated sequences (* denotes every10th residue).....	9
Figure S-2 B. Alignment of most abundant Ara h 1 isoform P43237 with detected peptides from this study.	11
Figure S-3 A. Alignment of Ara h 3 sequences used in this study (* denotes every10th residue).....	14
Figure S-3 B. Alignment of one abundant Ara h 3 isoform Q647H4_ARAHY with detected peptides from this study.....	17
Figure S-3 C. Alignment of a second abundant Ara h 3 isoform Q0GM57 with detected peptides from this study.....	19
Figure S-4 A. Alignment of Ara h 2 sequences used in this study from the compressed UniProt accession Q6PSU2 (* denotes every10th residue).....	21
Figure S-4 B. Alignment of an Ara h 2 isoform Q6PSU2 with detected peptides from this study.....	22
Figure S-5 A. Alignment of Ara h 6 sequences used in this study (* denotes every10th residue).....	23
Figure S-5 B. Alignment of an Ara h 6 isoform Q647G9 with detected peptides from this study.....	24
Table S-3. Peanut allergens profiled in three samples of raw peanuts and two batches roasted peanut flour using DIA-IM-MS.	25
Figure S-6 EIC for protein isoforms which were identified using a single peptide ID	31
Table S-4. Peanut proteins (other than allergens) identified in three samples of raw peanuts and two batches roasted peanut flour using DIA-IM-MS.....	33

Supplementary Materials and Methods

Reagents. Urea, thiourea, 3-[(3-cholamidopropyl) dimethylammonio]-2-hydroxy-1-propanesulfonate (CHAPS), dithiothreitol (DTT), were purchased from Sigma (Poole, Dorset, UK). Rapigest™ (acid-labile detergent) was obtained from Waters Corporation (Milford, MA, USA), 2D Clean-up Kit™, 2D Quant-Kit™, were supplied by GE Healthcare (Buckinghamshire, UK). Pre-cast 4-12 % BisTris gradient gels, MES(2-(*N*-morpholino)ethanesulfonic acid) buffer, SimplyBlue™ SafeStain, SeeBlue® pre-stained molecular weight markers including insulin β -chain (Mr 3 kDa), aprotinin (Mr 6 kDa), lysozyme (Mr 14 kDa), myoglobin (Mr 18 kDa), carbonic anhydrase (Mr 28 kDa), alcohol dehydrogenase (Mr 38 kDa), glutamic dehydrogenase (Mr 49 kDa), BSA (Mr 62 kDa), myosin (Mr 188 kDa), were purchased from Invitrogen (Paisley, UK).

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Peanut flour samples were prepared for SDS-PAGE by addition of NuPAGE LDS sample buffer (1 volume extract : 4 volumes LDS sample buffer) containing 50 mM DTT followed by heating at 80°C for 10 min. Samples were then centrifuged at 10,000 x g (5 min) prior to gel loading. SDS-PAGE analysis was undertaken using 4-12% Bis-Tris precast gradient gels (Invitrogen), 1.0 mm thickness and were subjected to electrophoresis using NuPage MES-SDS running buffer system under reducing conditions at a constant voltage of 200 V for 40 min. Gels were stained for 5-3 h using a Coomassie-based stain (Simply Blue Safe Stain, Invitrogen, Paisley, UK). Destaining was performed in distilled water. Gels were imaged at high resolution using a GE Healthcare Typhoon TRIO variable mode imager (Invitrogen, Paisley, UK).

Immunoblotting. Protein extracts were separated by SDS-PAGE and gels blotted onto nitrocellulose using a Biorad semi-dry blotting apparatus (Bio-Rad Laboratories, Hercules, CA, USA) using a transfer voltage of 15V for 20 min. Blots were blocked by incubation in phosphate buffered saline (137mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄, 1.8mM KH₂PO₄, pH 7.4 containing 0.5 % (v/v) Tween-20; PBS-T) and 5 % (w/v) skimmed milk powder overnight at 4°C. After washing in PBS-T blots were incubated with anti-peanut allergen polyclonal antibodies diluted 1:50,000 (v/v) in antibody dilution buffer (PBS-T containing 2.5% (w/v) skimmed milk powder) overnight at 4°C. Blots were washed in PBS-T before incubation with secondary antibody (goat-anti rabbit alkaline phosphatase linked IgG, Sigma Chemical Co, Poole, UK) diluted 1:10,000 (v/v) in antibody dilution buffer for 1 h. Blots were again washed in PBS-T before development using 5-bromo-4-chloro-3-indolyl phosphate (BCIP)- nitroblue tetrazolium (NBT) premade substrate solution (Thermo Scientific, Rockford, USA).

Mass spectrometry methods. Mobile phase A consisted of water containing 0.1% (v/v) formic acid, whilst mobile phase B was 100 % (v/v) acetonitrile containing 0.1% (v/v) formic acid. In preparation for LC/MS analyses, samples were transferred to total recovery vials and diluted with mobile phase A providing a working solution of 100 ng/ μ L with 1 μ L injected on-column. The sample was typically transferred with aqueous 0.1% (v/v) formic acid to the precolumn at a flow rate of 5 μ L/min for 3 min. After desalting and pre-concentration, peptides were eluted from the pre-column to the analytical column and separated using a gradient of 3-40 % mobile phase B over 90 min at a flow rate of 300 nL/min, followed by a 10 min column rinse with 85 % mobile phase B. Column re-equilibration to initial conditions occurred over 20 min. The analytical column temperature was maintained at 35 °C. The lock mass compound, [Glu¹]-Fibrinopeptide B (50 fmol/ μ L), was delivered by the auxiliary pump of the nanoAcquity system at 500 nL/min to the reference sprayer of the NanoLockSpray source of the mass spectrometer. Mass spectrometric analysis utilized a Synapt G2-S HDMS mass spectrometer (Waters Corporation, Manchester, UK). For all measurements, the mass spectrometer was operated in positive ion ESI, resolution mode with a nominal resolution of 20,000 FWHM. The ion source block temperature and capillary voltage were set to 70°C and 3.2 kV, respectively. The time-of-flight analyser (TOF) of the mass spectrometer was externally calibrated with a NaCsI mixture from *m/z* 50 to 1990. Data were post-acquisition lock mass corrected using the doubly charged monoisotopic ion of [Glu¹]-Fibrinopeptide B. The reference sprayer was sampled with a frequency of 60 s. Accurate mass LC-MS data were collected in data independent (DIA-IM-MS) mode of acquisition coupled with on-line ion mobility (IM). In low energy MS mode, data were collected at constant collision energy of 4 eV (per unit charge). In the elevated energy mode, the collision energy was ramped from 20 eV to 45 eV (per unit charge) during each integration. The spectral acquisition time in each mode was 0.49 s with a 0.01 s interscan delay. One cycle of low and elevated energy data was acquired every 1 s. The quadrupole mass analyser was operated in non-resolving RF only mode such that ions from *m/z* 300 to 2000 were efficiently transmitted, ensuring that any ions observed in the LC-MS data less than *m/z* 300 were known to arise from collision induced dissociation in the collision cell. The IM separation capability of the instrument has been described previously¹.

Peanut allergen database preparation and mass spectrometry data analysis. A curated peanut allergen sequence data set was generated using exemplar sequences obtained from the IUIS allergen nomenclature database (<http://www.allergen.org/>)². Homologues of each of these sequences were identified from the UniProt database using BLASTp³ against *A. hypogaea* protein sequences, with sequences of >70 % identity and appropriate protein description taken as allergen homologues. These sequences were aligned using ClustalW2⁴ using default settings. Dendrograms (average distance, BLOSUM62) were drawn using Jalview ver 2.7⁵. Dendrograms were presented and

annotated using Figtree ver. 1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree/>). An additional database comprising all known peanut protein accessions in UniProt Viridiplantae was used to identify all non-allergenic peanut proteins.

In order to detect and monitor protein and peptide identification error rates, decoy database entries were created on the fly with every search conducted by reversing the original protein amino acid sequences and concatenating them with the original database entries, FDR value was 4%. The principle of the database search algorithm used has been described previously⁶. The following criteria were used for protein identification: peptide and fragment ion tolerances (10 ppm) were determined automatically, one missed tryptic cleavage site allowed, fixed modification; carbamidomethylation of cysteine and variable modifications, oxidation of methionine and hydroxylation of proline.

Figure S-1. SDS-PAGE analysis and immunoblots of raw and roasted peanut flour extracts prepared using two different buffers.

Raw sample 2 (lanes 1, 5), raw sample 3 (blanched) (2, 6), raw sample 1 (3, 7) and roasted (4, 8) peanut flour was extracted using either 50 mM Tris-HCl pH 8.8, 50 mM DTT, 0.04 % (w/v) acid-labile detergent (lanes 1-4) or 7 M urea, 2 M thiourea, 2 % (w/v) CHAPS, 1 % (w/v) DTT (lanes 5-8). 5 mg of protein was loaded per sample lane. Protein stained SDS-PAGE shown in A; Western blots performed with peanut specific antibodies raised in rabbit to; 2S (B), 7S (C) and 11S (D) protein families. The relative mobility's of the major peanut allergens are indicated on the right.

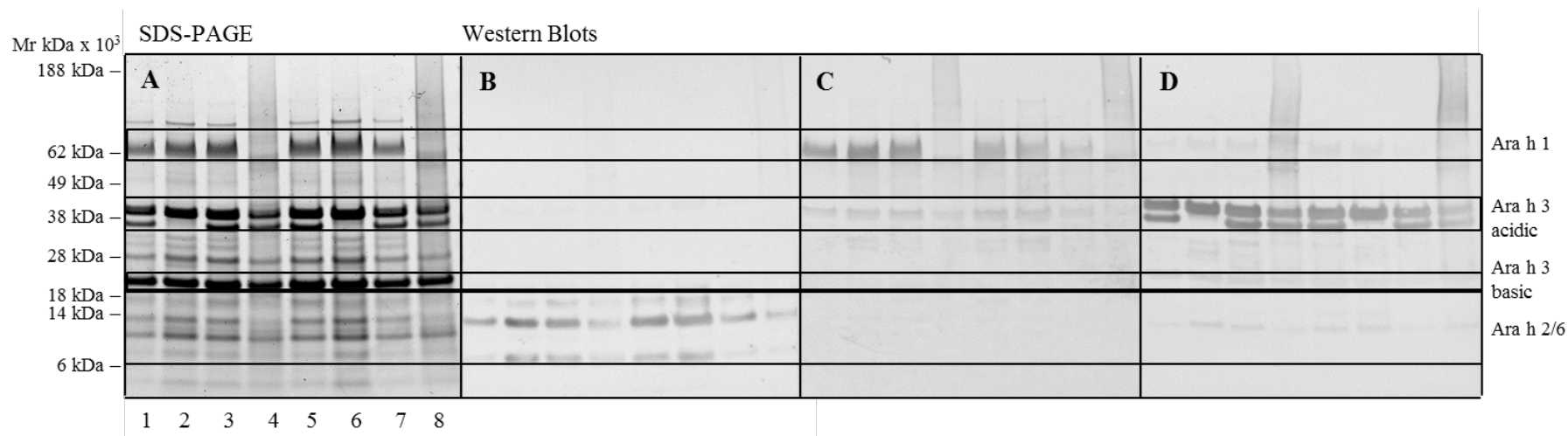


Table S-1. Effect of roasting on peanut protein extraction efficiency using two different buffers.

Peanut flour protein content was estimated to be between 41.4-44.6% (w/w) for the raw and 54.9 % (w/w) for the roasted peanut flour using a micro Kjeldahl method²⁷. The protein content of extracts was determined using the 2D-Quant assay and the extraction efficiency calculated as a percentage of the flour protein content determined by the Kjeldahl method. All extractions were performed in triplicate and are displayed as mean +/- SD, roasted peanut flour extractions were performed on two batches.

Extraction buffer	Raw peanut - 1		Raw peanut – 2		Raw peanut – 3 Blanched		Roasted peanut	
	Protein (mg/g)	% extraction	Protein (mg/g)	% extraction	Protein (mg/g)	% extraction	Protein (mg/g)	% extraction
50 mM Tris-HCl pH 8.8, 50 mM DTT, 0.04 % acid-labile detergent	9.00 +/- 0.8	108.3 +/- 6	9.00 +/- 0.8	108.3 +/- 6	8.61+/- 0.4	98.7 +/- 6	9.06+/-1.0	85.3+/-5
7 M urea, 2 M thiourea, 2 % (w/v) CHAPS, 1 % (w/v) DTT)	10.01 +/- 0.6	119.0 +/- 5	10.01 +/- 0.6	119.0 +/- 5	8.77 +/- 0.8	101.7 +/- 8	10.91+/-0.3	99.3+/-2

Scheme S-1. Workflow for the analysis of MS data using an initial screen of all peanut protein sequences (UniProt) and further analysis of selected allergen sequences.

Transomics latest version is now Progenesis QI for proteomics informatics

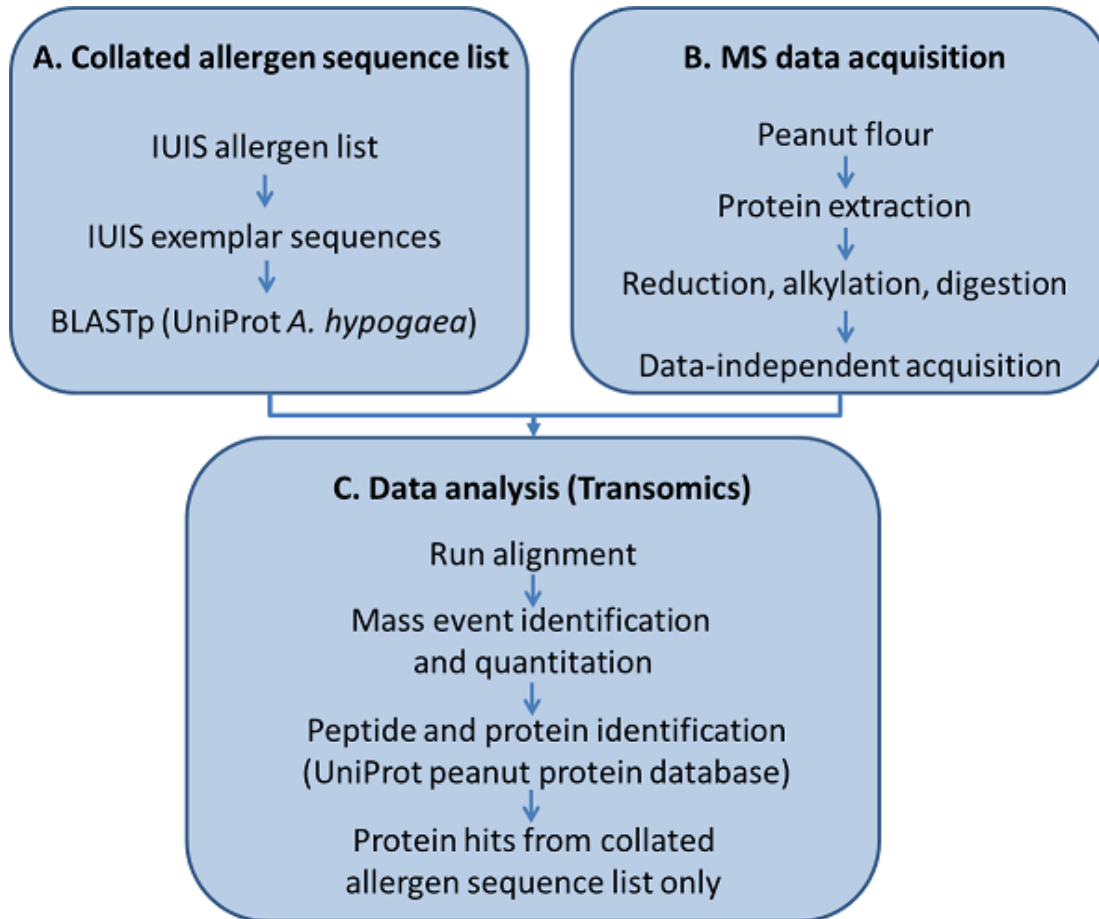


Table S-2. Additional information on the detection of the most abundant allergen isoforms in this study.

Allergen Designator	Uniprot Accession	Number of AA in mature seq (minus signal sequence)	Sum of aa in peptides identified	Percent of mature sequence identified	Theoretical Mr of mature sequence (average)	Theoretical pI of mature sequence
Ara h 1	P43237	537 (PRECURSOR not included)	352	65.55	61721.18	6.12
Ara h 3	Q647H4	518	339	65.44	59638.20	5.51
Ara h 3	Q0GM57	494	294	59.41	56415.90	5.35
Ara h 2	Q6PSU2	151	137	90.73	17993.59	5.51
Ara h 6	Q647G9	124	90	72.58	14845.61	5.49

Figure S-2 A. Alignment of Ara h 1 curated sequences (* denotes every10th residue).

```

      *      20      *      40      *      60      *      80      *      100
Q6PSU6_ : ~~~~~ : -
P43237_ : MRGRVSPIMLLLGILVLASVSATQAKS-PY-RKTENPCAQRCLQSCQQEPDDLKQKACESRCTKLEYDPRCVYD----TGATNQRHPPGERTRGRQPGDYDDDRRQ : 100
Q6PSU3_ : MRGRVSPIMLLLGILVLASVSATQAKS-PY-RKTENPCAQRCLQSCQQEPDDLKQKACESRCTKLEYDPRCVYD----TGATNQRHPPGERTRGRQPGDYDDDRRQ : 100
E5G076_ : MRGRVSPIMLLLGILVLASVSATHAKSSPYQKKTENPCAQRCLQSCQQEPDDLKQKACESRCTKLEYDPRCVYDPRGHTGTGTNQRSPPGERTRGRQPGDYDDDRRQ : 106
Q6PSU5_ : ~~~~~ : -
P43238_ : MRGRVSPIMLLLGILVLASVSATHAKSSPYQKKTENPCAQRCLQSCQQEPDDLKQKACESRCTKLEYDPRCVYDPRGHTGTGTNQRSPPGERTRGRQPGDYDDDRRQ : 106
Q6PSU4_ : ~~~~~ : -

      *      120      *      140      *      160      *      180      *      200      *
Q6PSU6_ : ~~~~~ : -
P43237_ : PRREEGGRWGPAEPREREREEDWRQPREDWRRPSSHQQPRKIRPEGREGEQEWTGPGSEVREETSRRNPFYFPSRRFSTRYGNQNGRIRVLQRFQDQSRQFQNLQNH : 206
Q6PSU3_ : PRREEGGRWGPAEPREREREEDWRQPREDWRRPSSHQQPRKIRPEGREGEQEWTGPGSEVREETSRRNPFYFPSRRFSTRYGNQNGRIRVLQRFQDQSRQFQNLQNH : 206
E5G076_ : PRREEGGRWGPAEPREREREEDWRQPREDWRRPSSHQQPRKIRPEGREGEQEWTGPGSHVREETSRRNPFYFPSRRFSTRYGNQNGRIRVLQRFQDQSRQFQNLQNH : 212
Q6PSU5_ : ~~~~~ : -
P43238_ : PRREEGGRWGPAEPREREREEDWRQPREDWRRPSSHQQPRKIRPEGREGEQEWTGPGSHVREETSRRNPFYFPSRRFSTRYGNQNGRIRVLQRFQDQSRQFQNLQNH : 212
Q6PSU4_ : ~~~~~ : ~~~~~GFDQSRQFQNLQNH : 15

      220      *      240      *      260      *      280      *      300      *      3
Q6PSU6_ : ~~~~~ : 1
P43237_ : RIVQIEARPNTLVLPKHADADNIIQGGQATVTVANGNNRKSFNLDDEGHALRIPSGFISYILNRHDNQNLRVAKISMPVNTPGQFEDFFPASSRDQSSYLQGFSR : 312
Q6PSU3_ : RIVQIEARPNTLVLPKHADADNIIQGGQATVTVANGNNRKSFNLDDEGHALRIPSGFISYILNRHDNQNLRVAKISMPVNTPGQFEDFFPASSRDQSSYLQGFSR : 312
E5G076_ : RIVQIEAKPNTLVLPKHADADNIIQGGQATVTVANGNNRRALILTRAMHS--ESHPPHFLHDDMTPELRVA-KSHAVNTPGQFEDFFPASSRDQSSYLQGFSR : 315
Q6PSU5_ : ~~~~~ : -
P43238_ : RIVQIEAKPNTLVLPKHADADNIIQGGQATVTVANGNNRKSFNLDDEGHALRIPSGFISYILNRHDNQNLRVAKISMPVNTPGQFEDFFPASSRDQSSYLQGFSR : 318
Q6PSU4_ : RIVQIEAKPNTLVLPKHADADNIIQGGQATVTVANGNNRKSFNLDDEGHALRIPSGFISYILNRHDNQNLRVAKISMPVNTPGQFEDFFPASSRDQSSYLQGFSR : 121

      20      *      340      *      360      *      380      *      400      *      420
Q6PSU6_ : NTLEAAAFNAEFNEIRRVLLEENAGGEQEERGQRRRSTRSS-DNEGVIVKVSKEHVQELTKHAKSVSKKGSEE-EDITNPINLRDGEPLDLSNNFGRLFVVKPDKKNP : 105
P43237_ : NTLEAAAFNAEFNEIRRVLLEENAGGEQEERGQRRRSTRSS-DNEGVIVKVSKEHVQELTKHAKSVSKKGSEE-EDITNPINLRDGEPLDLSNNFGRLFVVKPDKKNP : 416
Q6PSU3_ : NTLEAAAFNAEFNEIRRVLLEENAGGEQEERGQRRRSTRSS-DNEGVIVKVSKEHVQELTKHAKSVSKKGSEE-EDITNPINLRDGEPLDLSNNFGRLFVVKPDKKNP : 416
E5G076_ : NTLEAAAFNAEFNEIRRVLLEENAGGEQEERGQRRRSTRSSDNEGVIVKVSKEHVEELTKHAKSVSKKGSEEGDITNPINLRDGEPLDLSNFGRLFVVKPDKKNP : 421
Q6PSU5_ : ~LEAAAFNAEFNEIRRVLLEENAGGEQEERGQRRRSTRSSDNEGVIVKVSKEHVEELTKHAKSVSKKGSEEGDITNPINLRDGEPLDLSNFGRLFVVKPDKKNP : 104
P43238_ : NTLEAAAFNAEFNEIRRVLLEENAGGEQEERGQRRRSTRSSDNEGVIVKVSKEHVEELTKHAKSVSKKGSEEGDITNPINLRDGEPLDLSNFGRLFVVKPDKKNP : 424
Q6PSU4_ : NTLEAAAFNAEFNEIRRVLLEENAGGEQEERGQRRRSTRSSDNEGVIVKVSKEHVEELTKHAKSVSKKGSE-EGDITNPINLRDGEPLDLSNFGRLFVVKPDKKNP : 226

```

		*	440	*	460	*	480	*	500	*	520	*	
Q6PSU6_	:	QLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVVNKG	TGNLELVAVRKEQQQRGRREQEWE	EEEE	EEEEEGSNREVRRYTARLKEGDVFIMPAAHPVAINASSELH	:	211						
P43237_	:	QLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVVNKG	TGNLELVAVRKEQQQRGRREQEWE	EEEE	EEEEEGSNREVRRYTARLKEGDVFIMPAAHPVAINASSELH	:	522						
Q6PSU3_	:	QLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVVNKG	TGNLELVAVRKEQQQRGRREQEWE	EEEE	EEEEEGSNREVRRYTARLKEGDVFIMPAAHPVAINASSELH	:	522						
E5G076_	:	QLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVINKG	TGNLELVAVRKEQQQRGRREQEWE	EEEE	EEEEEGSNREVRRYTARLKEGDVFIMPAAHPVAINASSELH	:	527						
Q6PSU5_	:	QLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVVNKG	TGNLELVAVRKEQQQRGRREEE	---	EDEDQEEEGSNREVRRYTARLKEGDVFIMPAAHPVAINASSELH	:	207						
P43238_	:	QLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVVNKG	TGNLELVAVRKEQQQRGRREEE	---	EDEDQEEEGSNREVRRYTARLKEGDVFIMPAAHPVAINASSELH	:	527						
Q6PSU4_	:	QLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVVNKG	TGNLELVAVRKEQQQRGRREEE	---	EDEDQEEEGSNREVRRYTARLKEGDVFIMPAAHPVAINASSELH	:	329						

		540	*	560	*	580	*	600	*	620	
Q6PSU6_	:	LLGFGINAENNHRIFLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKLIKQRESHFVSAR	PQSQSP	-----	SSPEKEDQEEENQGGKGPLL	SILKAFN	:	303			
P43237_	:	LLGFGINAENNHRIFLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKLIKQRESHFVSAR	PQSQSP	-----	SSPEKEDQEEENQGGKGPLL	SILKAFN	:	614			
Q6PSU3_	:	LLGFGINAENNHRIFLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKLIKQRESHFVSAR	~~~~~	~~~~~	~~~~~	~~~~~	:	580			
E5G076_	:	LLGFGINAENNHRIFLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKLIKQRESHFVSAR	PQSQSP	-----	SSPEKEDQEEENQGGKGPLL	SILKAFN	:	619			
Q6PSU5_	:	LLGFGINAENNHRIFLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKLIKQRESHFVSAR	PQSQSP	-----	SSPEKEDQEEENQGGKGPLL	SILKAFN	:	299			
P43238_	:	LLGFGINAENNHRIFLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKLIKQKESHFVSAR	PQSQSQSPSSPEKE	SSPEKEDQEEENQGGKGPLL	SILKAFN	:	626				
Q6PSU4_	:	LLGFGINAENNHRIFLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKLIKQKESHFVSAR	PQSQSQSPSSPEKE	SSPEKEDQEEENQGGKGPLL	SILKAFN	:	428				

Figure S-2 B. Alignment of most abundant Ara h 1 isoform P43237 with detected peptides from this study.

The signal sequence is underlined in italics; the precursor sequence in bold italics and the main sequence in bold and shaded where peptides were identified. Roasted peptides identified by DIA-IM-MS are aligned with the sequence, the most abundant peptides are underlined, raw peptides are in italics with the site of hydroxylation bold (P) and peptides found in both raw and roasted are in bold. The N-terminal sequences identified in purified Ara h 1 are marked (*)⁷ and the N-glycosylation site is boxed.

SIGNAL SEQUENCE MRGRVSPLMLLLGILVLASVSA

PRECURSOR SEQUENCE
TQAKSPYRK*TENPCAQRCLQSCQQEPDDLKQKACESRCTKLEYDPR**CVYDTGATN***

SPYRK CLQSCQQEPDDLK LEYDPR
CLQSCQQEPDDLKQK
CLQSCQQEPDDLKQKACESR
KTENPCAQR ACESRCTKLEYDPR
KTENPCAQR CTKLEYDPR
SPYRK TENPCAQR CTKLEYDPR
TENPCAQR LEYDPRCVYDTGATN
CVYDTGATN
LEYDPRCVYDTGATN

MATURE SEQUENCE

QRHPPG*TRGRQPGDYDDRRQPRREEGGRWGPAEPREREREEDWRQ**PREDWRR*PSHQPRKIRPEGREGEQEWGTPGSEVR**

GRQPGDYDDRR WGPAEPREREREEDWR EDWRRPSHQQPR EGEQEWGTPGSEVR
GRQPGDYDDRR WGPAERER EEDWRQPR RPSHQQPR EGEQEWGTPGSEVR
GRQPGDYDDRR WGPAEPR EREDWR EDWRRPSHQQPR EGEQEWGTPGSEVR
QPGDYDDRRQPR EEGGRWGPAEPR EEDWRQPR EDWR EGEQEWGTPGSEVR
TRGRQPGDYDDRR REEGGRWGPAEPR EEDWRQPR EDWR EGEQEWGTPGSEVR
QRHPPGTR QPGDYDDRR EEGGRWGPAEPRERER QPR EGEQEWGTPGSEVR
QRHPPGER QPGDYDDRR EEGGRWGPAEPRER QPR EGEQEWGTPGSEVR
HPPGTR WGPAEPREREREEDWR IRPEGR
HPPGER QPRREEGGR EEDWR KIRPEGR
QPRREEGGR QPR QPR QPR QPR
REEGGRWGPAEPR
WGPAEPRERER

EETSRNNPFYFPSRRFSTRYGNQNGRIRVLQRFQDQSKQFQNLQNHRIQVIEARPNTLVLPKHADADNILVIQQGQATVTVANGNNRK
 EETSR YGNQNGR VLQRRDQR IVQIEARPNTLVLPK K
EETSRNNPFYFPSR SKQFQNLQNHR HADADNILVIQQGQATVTVANGNNRK
 EETSRNNPFYFPSRR QFQNLQNHR HADADNILVIQQGQATVTVANGNNR
 EETSRNNPFYFPSR
 EETSR
 EETSRNNPFYFPSR
 NNPFFYFPSRR
 NNPFFYFPSRR
 NNPFFYFPSRR
 NNPFFYFPSR
 NNPFFYFPSR
 NNPFFYFPSR
 NNPFFYFPSRRFSTR

SFNLDEGHALRIPSGFISYILNRHDNQNLRVAKISMPVNTPGQFEDFFPASSRDQSSYLQGFSRNTLEAAFNAEFNEIRR
 SFNLDEGHALR HDNQNLR ISMPVNTPGQFEDFFPASSR NTLEAAFNAEFNEIRR
 SFNLDEGHALRIPSGFISYILNRHDNQNLR ISMPVNTPGQFEDFFPASSR NTLEAAFNAEFNEIR
 SFNLDEGHALRIPSGFISYILNRHDNQNLR DQSSYLQGFSR
SFNLDEGHALR VAKISMPVNTPGQFEDFFPASSRDQSSYLQGFSR
 IPSGFISYILNR VAKISMPVNTPGQFEDFFPASSR
 IPSGFISYILNRHDNQNLR

VLLEENAGGEQEERGQRRRSTRSSDNEGVIVKVSKEHVQELTKHAKSVSKKGSEEDITNPINLRDGEPLSNNFGR
 VLLEENAGGEQEER SSDNEGVIVKVS GSEEDITNPINLRDGEPLSNNFGR
VLLEENAGGEQEER SSDNEGVIVK EHVQELTK GSEEDITNPINLRDGEPLSNNFGR
 VLLEENAGGEQEERGQR STRSSDNEGVIVK GSEEDITNPINLRDGEPLSNNFGR
 DGEPLSNNFGR
GSEEDITNPINLR
 GSEEDITNPINLR
 KGSEEDITNPINLR
 KGSEEDITNPINLR
 KGSEEDITNPINLRDGEPLSNNFGR
 SVSKKGSEEDITNPINLR

LFEVKPDKKNPQLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVVNKG TGNLELVAVRKEQQQRGRREQEWEEEEEEEEEGSNREVR RYTARLKEGDVFIMPAHP			
LFEVKPDKK	<u>EGALMLPHFNSK</u>	<u>GTGNLELVAVR</u>	EQEWEEEEEEEEEGSNREVR
PDKK		<i>GTGNLELVAVRK</i>	REQEWEEEEEEEEEGSNR
<i>LFEVKPDKKNPQLQDLDMMLTCVEIK</i>	<u>AMVIVVVNK</u>	GNLELVAVR	<i>EQEWEEEEEEEEEGSNR</i>
<u>LFEVKPDK</u>			<i>EQEWEEEEEEEEEGSNREVRR</i>
NPQLQDLDMMLTCVEIK			
<i>NPQLQDLDMMLTCVEIK</i>			
<i>KNPQLQDLDMMLTCVEIK</i>			
<i>KNPQLQDLDMMLTCVEIK</i>			
<i>NPQLQDLDMMLTCVEIKEGALMLPHFNSK</i>			

VAI <u>NAS</u> SELHLLGFGINAENNHRI FLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKLIK NQRESHFVSARPQSQSPSSPEKEDQEEENQGGKGPLL SILKAFN		
<u>IFLAGDKDNVIDQIEK</u>	<u>DLAFPGSGEQVEK</u>	<i>ESHFVSARPQSQSPSSPEK</i>
IFLAGDK	DLAFPGSGEQVEKLIK	<i>ESHFVSARPQSQSPSSPEK</i>
DNVIDQIEK	<i>DLAFPGSGEQVEK</i>	<i>EDQEEENQGGKGPLL</i> SILK
<i>DNVIDQIEKQAK</i>		
	<i>QAKDLAFPGSGEQVEK</i>	
<i>IFLAGDKDNVIDQIEKQAK</i>		

Figure S-3 A. Alignment of Ara h 3 sequences used in this study (* denotes every 10th residue).

		*	20	*	40	*	60	*	80	*	100	
Q647H2_	:	~MAKLI	ALSVCFCFLVLGASSVTFRQ	QGEENE	CQFQRLNAQRPDNR	IESEGGYIETWNPNNQ	EFQCAGVALSR	VLRRNALRRPFYSNAP	QEIFIY	QGS	GYFGLIF	: 105
E5G077_	:	~MAKLI	ALSVCFCFLVLGASSVTFRQ	QGEENE	CQFQRLNAQRPDNR	IESEGGYIETWNPNNQ	EFQCAGVALSR	VLRRNALRRPFYSNAP	LEIYV	QGS	GYFGLIF	: 105
Q6IWG5_	:	~KLI	ALSVCFCFLVLGASSVTFRQ	QGEENE	CQFQRLNAQRPDNR	IESEGGYIETWNPNNQ	EFQCAGVALSR	VLRRNALRRPFYSNAP	LEIYV	QGS	GYFGLIF	: 103
Q0GM57_	:	~MAKLI	ALSVCFCFLVLGASSVTFRQ	QGEENE	CQFQRLNAQRPDNR	IESEGGYIETWNPNNQ	EFQCAGVALSR	VLRRNALRRPFYSNAP	LEIYV	QGS	GYFGLIF	: 105
Q6T2T4_	:	~MGKLI	ALSVCFCFLVLGASSISFRQ	QPEENACQFQRLNAQRPDNR	IESEGGYIETWNPNNQ	EFECAGVALSR	VLRRNALRRPFYSNAP	QEIFIQ	QGR	AYFGLIF	: 105	
Q647H4_	:	~MGKLI	ALSVCFCFLVLGASSISFRQ	QPEENACQFQRLNAQRPDNR	IESEGGYIETWNPNNQ	EFECAGVALSR	VLRRNALRRPFYSNAP	QEIFIQ	QGR	GYFGLIF	: 105	
Q8LKN1_	:	~MGKLI	ALSVCFCFLVLGASSISFRQ	QPEENACQFQRLNAQRPDNR	IESEGGYIETWNPNNQ	EFECAGVALSR	VLRRNALRRPFYSNAP	QEIFIQ	QGR	GYFGLIF	: 105	
Q9FZ11_	:	MIRGR	IALSVCFCLVLGASSISFRQ	QPEENACQFQRLNAQRPDNR	LESEGGYIETWNPNNQ	EFECAGVALSR	VLRRNALRRPFYSNAP	QEIFIQ	QGR	GYFGLIF	: 106	
Q647H3_	:	~MAKLI	ALSVCFCFLVLGASSISFRQ	QPEENACQFQRLNAQRPDNR	LESEGGYIETWNPNNQ	EFECAGVALSR	VLRRNALRRPFYSNAP	QEIFIQ	QGR	GYFGLIF	: 105	
Q9SQH7_	:	~MAKLI	ELSF	CFCLVLGASSISFRQ	QPEENACQFQRLNAQRPDNR	IESEGGYIETWNPNNQ	EFECAGVALSR	VLRRNALRRPFYSNAP	QEIFIQ	QGR	GYFGLIF	: 105
A1DZF0_	:	~MAKLI	ELSF	CFCLVLGASSISFRQ	QPEENACQFQRLNAQRPDNR	IESEGGYIETWNPNNQ	EFECAGVALSR	VLRRNALRRPFYSNAP	QEIFIQ	QGR	GYFGLIF	: 105
O82580_	:	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	: 82
Q5I6T2_	:	~MAKLI	ELSF	CFCLVLGASSISFRQ	QPEENACQFQRLNAQRPDNR	IESEGGYIETWNPNNQ	EFECAGVALSR	VLRRNALRRPFYSNAP	QEIFIQ	QGR	GYFGLIF	: 105
B5TYU1_	:	~MAKLI	ELSF	CFCLVLGASSISFRQ	QPEENACQFQRLNAQRPDNR	IESEGGYIETWNPNNQ	EFECAGVALSR	VLRRNALRRPFYSNAP	QEIFIQ	QGR	GYFGLIF	: 105

		*	120	*	140	*	160	*	180	*	200	*			
Q647H2_	:	PGCP	GTFEEPIQ	GSEQF--	QRESRHEFC	GQDQS--	CRPIDTHQKVHGER	EGDLIAVEH	GVAFWIYND	QD	TDVVAISV	LHTNSLHN	QLDQFPRRFNLAGKQEQEFLRY	: 207	
E5G077_	:	PGCP	STYEPAQ	EGRRYQS	QKPSRRFC	VGQDDPS	CCQD	SHQKVHRF	DEGDLIAVPT	GVAFWMYN	EDTDVVT	VTLS	DTSSIHNQLDQFPRRFYLAGNQEQEFLRY	: 211	
Q6IWG5_	:	PGCP	STYEPAQ	EGRRYQS	QKPSRRFC	VGQDDPS	CCQD	SHQKVHRF	DEGDLIAVPT	GVAFWMYN	EDTDVVT	VTLS	DTSSIHNQLDQFPRRFYLAGNQEQEFLRY	: 209	
Q0GM57_	:	PGCP	STYEPAQ	EGRRYQS	QKPSRRFC	VGQDDPS	CCQD	SHQKVHRF	DEGDLIAVPT	GVAFWMYN	EDTDVVT	VTLS	DTSSIHNQLDQFPRRFYLAGNQEQEFLRY	: 211	
Q6T2T4_	:	IGCP	STYEPAQ	QGRRHQS	QRPERREC	GQDQSCC	QD	SHQKVHRF	DEGDLIAVPT	GVAFWMYN	EDTDVVA	VS	LTDTNNNDNQLDQFPRRFNLAGNHEQEFLRY	: 209	
Q647H4_	:	PGCP	STYEPAQ	QGRRHQS	QRPERREC	GQDQSCC	QD	SHQKVHRF	DEGDLIAVPT	GVAFWMYN	EDTDVVA	VS	LTDTNNNDNQLDQFPRRFNLAGNHEQEFLRY	: 209	
Q8LKN1_	:	PGCP	STYEPAQ	QGRRHQS	QRPERREC	GQDQSCC	QD	SHQKVHRF	DEGDLIAVPT	GVAFWMYN	EDTDVVA	VS	LTDTNNNDNQLDQFPRRFNLAGNHEQEFLRY	: 209	
Q9FZ11_	:	PGCP	STYEPAQ	QGRRHQS	QRAPRREE	GEDQSCC	QD	SHQKVRRF	DEGDLIAVPT	GVAIWMF	NDH	DTDVVA	VS	LTDTNNNDNQLDQFPRRFNLAGNHEQEFLRY	: 211
Q647H3_	:	PGCP	STYEPAQ	QGRRHQS	QRAPRREE	GEDQSCC	QD	SHQKVRRF	DEGDLIAVPT	GVAIWMY	NDH	DTDVVA	VS	LTDTNNNDNQLDQFPRRFNLAGNHEQEFLRY	: 210
Q9SQH7_	:	PGCP	STYEPAQ	QGRRYQS	QRPERRLC	EEDQSCC	QD	SHQKVHRF	NEGDLIAVPT	GVAFWLYN	EDTDVVA	VS	LTDTNNNDNQLDQFPRRFNLAGNHEQEFLRY	: 209	
A1DZF0_	:	PGCP	STYEPAQ	QGRRYQS	QRPERRLC	EEDQSCC	QD	SHQKVHRF	NEGDLIAVPT	GVAFWLYN	EDTDVVA	VS	LTDTNNNDNQLDQFPRRFNLAGNHEQEFLRY	: 209	
O82580_	:	PGC	ERHYEE	PHTQ	GRRSQR	PERRLC	GEDQSCC	QD	SHQKVHRF	DEGDLIAVPT	GVAFWLYN	EDTDVVA	VS	LTDTNNNDNQLDQFPRRFNLAGNTEQEFLRY	: 186
Q5I6T2_	:	PGCP	STYEPAQ	QGRRSQR	PERRLC	GEDQSCC	QD	SHQKVHRF	DEGDLIAVPT	GVAFWLYN	EDTDVVA	VS	LTDTNNNDNQLDQFPRRFNLAGNHEQEFLRY	: 209	
B5TYU1_	:	PGCP	STYEPAQ	QGRRYQS	QRPERRLC	EEDQSCC	QD	SHQKVHRF	NEGDLIAVPT	GVAFWLYN	EDTDVVA	VS	LTDTNNNDNQLDQFPRRFNLAGNHEQEFLRY	: 209	

	220	*	240	*	260	*	280	*	300	*	3		
Q647H2_	QQRSGRQS				PKG		EEQE	EQENEGGNVFSGFST	EFLSHGFQVN	EDIVRNLRGENERE	EQGAIVTVKGGLS	SILVPEWRQ	: 284
E5G077_	QQCQGSRP	Y		RQISP		RVRG	DEQENEGSNIFSGFAQ	EFLQHA	FQVD	RQTVENLRGENERE	EQGAIVTVKGGLR	ILSPDEEDE	: 292
Q6IWG5_	QQCQGSRP	Y		RQISP		RVRG	DEQENEGSNIFSGFAQ	EFLQHA	FQVD	RQTVENLRGENERE	EQGAIVTVKGGLR	ILSPDEEDE	: 290
Q0GM57_	QQCQGSRP	Y		RQISP		RVRG	DEQENEGSNIFSGFAQ	EFLQHA	FQVD	RQTVENLRGENERE	EQGAIVTVKGGLR	ILSPDEEDE	: 292
Q6T2T4_	QQC		SRRSLPSPYSPQTQPKQ	EDREFSPRGQH	RRERAGQE	EQENEGGNIFSGFT	PEFLAQA	FQVDDRQILQ	NLRGENES	DEQGAIVTVRGGLR	ILSPDRKRR		: 312
Q647H4_	QQC		SRRSLPSPYSPQTQPKQ	EDREFSPRGQH	RRERAGQE	EQENEGGNIFSGFT	PEFLAQA	FQVDDRQILQ	NLRGENES	DEQGAIVTVRGGLR	ILSPDRKRR		: 312
Q8LKN1_	QQC		SRRSLPSPYSPQTQPKQ	EDREFSPRGQH	RRERAGQE	EQENEGGNIFSGFT	PEFLAQA	FQVDDRQILQ	NLRGENES	DEQGAIVTVRGGLR	ILSPDRKRR		: 312
Q9FZ11_	QQC		SRRSLPSPYSPQSQPRQE	ERFSPRGQHS	RRERAGQEE	ENEGGNIFSGFT	PEFLAQA	FQVDDRQIVQ	NLRGENES	EEGAIVTVKGGLR	ILSPDRKRG		: 314
Q647H3_	QQC		SRRSLPSPYSPQSQPRQE	ERFSPRGQHS	RRERAGQE	EQENEGGNIFSGFT	PEFLAQA	FQVDDRQILQ	NLRGENES	DEQGAIVTVRGGLR	ILSPDRKRR		: 313
Q9SQH7_	QQCSRQ		SRRSLPSPYSPHSRPR	REERFSPRGQHS	RRERAGQEE	DEGGNIFSGFT	PEFLAQA	FQVDDRQIVQ	NLRGENES	EEGAIVTVRGGLR	ILSPDRKRG		: 315
A1DZF0_	QQCSRQ		SRRSLPSPYSPQ	PGQEDREFSPRGQH	RRERAGQE	EQENEGGNIFSGFT	SEFLAQA	FQVDDRQIVQ	NLRGENES	EEGAIVTVKGGLR	ILSPDRKS		: 312
O82580_	QQCSRQ		SRRSLPSPYSPQSQPRQE	ERFSPRGQHS	RRERAGQEE	ENEGGNIFSGFT	PEFLAQA	FQVDDRQIVQ	NLRGENES	EEGAIVTVRGGLR	ILSPDRKRR		: 292
Q5I6T2_	QQCSRQ		SRRSLPSPYSPQSQPRQE	ERFSPRGQHS	RRERAGQEE	ENEGGNIFSGFT	PEFLAQA	FQVDDRQIVQ	NLRGENES	EEGAIVTVRGGLR	ILSPDRKRG		: 315
B5TYU1_	QQCSRQ		SRRSLPSPYSPQSQPRQE	ERFSPRGQHS	RRERAGQEE	ENEGGNIFSGFT	PEFLAQA	FQVDDRQIVQ	NLRGENES	EEGAIVTVRGGLR	ILSPDRKRG		: 315

	20	*	340	*	360	*	380	*	400	*	420	
Q647H2_	S				YQQPGRGDKDFNNGIEETICTATVKM	NIGKST	SADIYNPQAGSVRTV	NEL		DLFILNRLGLSAEYGS	SIHRDAMFVPHYN	: 363
E5G077_	SSRSPPNRR	EEFDEDRSRP			QQRGKYDENRRGYKNGIEETICSASV	VKKNLGRSSNP	DIYNPQAGSLRSV	NEL		DLFILGLWLGLSAQHGTIYRNAMFVPHYT		: 391
Q6IWG5_	SSRSPPSR	EEFDEDRSRP			QQRGKYDENRRGYKNGIEETICSASV	VKKNLGRSSNP	DIYNPQAGSLRSV	NEL		DLFILGLWLGLSAQHGTIYRNAMFVPHYT		: 389
Q0GM57_	SSRSPPSR	EEFDEDRSRP			QQRGKYDENRRGYKNGIEETICSASV	VKKNLGRSSNP	DIYNPQAGSLRSV	NEL		DLFILGLWLGLSAQHGTIYRNAMFVPHYT		: 391
Q6T2T4_	QQYERP	DEEEYDEDEYDEEERQ		QDRRRGRGSRGSGNGIEETICTAS	FKKNIGRNRSP	DIYNPQAGSLKTAN	EL		NLLILRWLGLSAEYGNLYRNALFVPHYN		: 416	
Q647H4_	QQYERP	DEEEYDEDEYDEEERQ		QDRRRGRGSRGSGNGIEETICTAS	FKKNIGRNRSP	DIYNPQAGSLKTAN	EL		NLLILRWLGLSAEYGNLYRNALFVPHYN		: 416	
Q8LKN1_	QQYERP	DEEEYDEDEYDEEERQ		QDRRRGRGSRGSGNGIEETICTAS	FKKNIGRNRSP	DIYNPQAGSLKTAN	EL		NLLILRWLGLSAEYGNLYRNALFVPHYN		: 418	
Q9FZ11_		ADDEEEYDEDEYDEE		DRRRGRGSRGSGNGIEETICTASV	VKKNIGRNRSP	DIYNPQAGSLKTAN	DL		NLLILRWLGLSAEYGNLYRNALFVPHYN		: 409	
Q647H3_	QQYERP	DEEEYDEDEYDEEERQ		QDRRRGRGSRGSGNGIEETICTASV	VKKNIGRNRSP	DIYNPQAGSLKTAN	DL		NLLILRWLGLSAEYGNLYRNALFVPHYN		: 417	
Q9SQH7_		ADDEEEYDEDEYDEE		QDGRGRGSRGSGNGIEETICTACV	VKKNIGRNRSP	HIYDPQRWFTQ	NCHDL		NLLILRWLGLSAEYGNLYRNALFVPHYN		: 410	
A1DZF0_		PDDEEEYDEDEYAEERQ		QDRRRGRGSRGSGNGIEETICTATV	VKKNIGRNRSP	DIYNPQAGSLKTAN	EL		NLLILRWLGLSAEYGNLYRNALFVPHYN		: 409	
O82580_		ADDEEEYDEDEYDEE		EDRRGRGSRGSGNGIEETICTASAK	NIGRNRSP	DIYNPQAGSLKTAN	DL		NLLILRWLGLSAEYGNLYRNALFVPHYN		: 387	
Q5I6T2_		ADDEEEYDEDEYDEE		EDRRGRGSRGSGNGIEETICTASV	VKKNIGRNRSP	DIYNPQAGSLKTAN	DL		NLLILRWLGLSAEYGNLYRNALFVPHYN		: 410	
B5TYU1_		ADDEEEYDEDEYDEE		EDRRGRGSRGSGNGIEETICTATV	VKKNIGRNRSP	DIYNPQAGSLKTAN	EL		NLLILRWLGLSAEYGNLYRNALFVPHYN		: 410	

		*	440	*	460	*	480	*	500	*	520	*									
Q647H2_	:	MNANS	MIYALHGG	AHVQVVD	CNGNRV	FDEELQEG	QSLVVP	QNFAVA	AKSQSE	HELYVAF	KTNSRAS	ISNLAG	KNSYMW	NLPEDV	VANSYGL	QYEQAR	QLKNNNPF	:	468		
E5G077_	:	LNAHTI	VVALNG	RAHVQV	VDSNGNR	VYDEELQ	EGHVLV	VPQNFA	VAAKAQ	SENYEY	LAFKTDS	RPSIAN	QAGENS	IIDNL	PEEVVA	NSYRLP	REQAR	QLKNNNPF	:	496	
Q6IWG5_	:	LNAHTI	VVALNG	RAHVQV	VDSNGNR	VYDEELQ	EGHVLV	VPQNFA	VAAKAQ	SENYEY	LAFKTDS	RPSIAN	LAGENS	IIDNL	PEEVVA	NSYRLP	REQAR	QLKNNNPF	:	494	
Q0GM57_	:	LNAHTI	VVALNG	RAHVQV	VDSNGNR	VYDEELQ	EGHVLV	VPQNFA	VAAKAQ	SENYEY	LAFKTDS	RPSIAN	LAGENS	IIDNL	PEEVVA	NSYRLP	REQAR	QLKNNNPF	:	496	
Q6T2T4_	:	TNAHSI	IYALRG	RAHVQV	VDSNGDR	VFDEELQ	EGHVLV	VPQNFA	VAGKQ	SENFEE	YVAFKT	DSRPSI	ANLAG	ENSEI	IDNL	PEEVVA	NSYGLP	REQAR	QLKNNNPF	:	521
Q647H4_	:	TNAHSI	IYALRG	RAHVQV	VDSNGDR	VFDEELQ	EGHVLV	VPQNFA	VAGKQ	SENFEE	YVAFKT	DSRPSI	ANLAG	ENSEI	IDNL	PEEVVA	NSYGLP	REQAR	QLKNNNPF	:	521
Q8LKN1_	:	TNAHSI	IYALRG	RAHVQV	VDSNGDR	VFDEELQ	EGHVLV	VPQNFA	VAGKQ	SENFEE	YVAFKT	DSRPSI	ANLAG	ENSEI	IDNL	PEEVVA	NSYGLP	REQAR	QLKNNNPF	:	523
Q9FZ11_	:	TNAHSI	IYALRG	RAHVQV	VDSNGNR	VYDEELQ	EGHVLV	VPQNFA	VAGKQ	SDNFE	YVAFKT	DSRPSI	ANLAG	ENSEI	IDNL	PEEVVA	NSYGLP	REQAR	QLKNNNPF	:	514
Q647H3_	:	TNAHSI	IYALRG	RAHVQV	VDSNGNR	VYDEELQ	EGHVLV	VPQNFA	VAGKQ	SDNFE	YVAFKT	DSRPSI	ANLAG	ENSEI	IDNL	PEEVVA	NSYGLP	REQAR	QLKNNNPF	:	522
Q9SQH7_	:	TNAHSI	IYALRG	RAHVQV	VDSNGNR	VYDEELQ	EGHVLV	VPQNFA	VAGKQ	SENFEE	YVAFKT	DSRPSI	ANLAG	ENSEI	IDNL	PEEVVA	NSYGLP	REQAR	QLKNNNPF	:	515
A1DZF0_	:	TNAHSI	IYALRG	RAHVQV	VDSNGNR	VYDEELQ	EGHVLV	VPQNFA	VAGKQ	SENFEE	YVAFKT	DSRPSI	ANLAG	ENSEI	IDNL	PEEVVA	NSYGLP	REQAR	QLKNNNPF	:	514
O82580_	:	TNAHSI	IYRLRG	RAHVQV	VDSNGNR	VYDEELQ	EGHVLV	VPQNFA	VAGKQ	SENFEE	YVAFKT	DSRPSI	ANLAG	ENSEI	IDNL	PEEVVA	NSYGLP	REQAR	QLKNNNPF	:	492
Q5I6T2_	:	TNAHSI	IYALRG	RAHVQV	VDSNGNR	VYDEELQ	EGHVLV	VPQNFA	VAGKQ	SDNFE	YVAFKT	DSRPSI	ANLAG	ENSEI	IDNL	PEEVVA	NSYGLP	REQAR	QLKNNNPF	:	516
B5TYU1_	:	TNAHSI	IYALRG	RAHVQV	VDSNGNR	VYDEELQ	EGHVLV	VPQNFA	VAGKQ	SDNFE	YVAFKT	DSRPSI	ANLAG	ENSEI	IDNL	PEEVVA	NSYGLP	REQAR	QLKNNNPF	:	515

		540		
Q647H2_	:	TFLVPP	QDSQM	IRTVA : 484
E5G077_	:	KFFVPP	FDHQSM	REVA : 512
Q6IWG5_	:	KFFVPP	FDHQSM	REVA : 510
Q0GM57_	:	KFFVPP	FDHQSM	REVA : 512
Q6T2T4_	:	KFFVPP	SEQ-SL	RAVA : 536
Q647H4_	:	KFFVPP	SEQ-SL	RAVA : 536
Q8LKN1_	:	KFFVPP	SEQ-SL	RAVA : 538
Q9FZ11_	:	KFFVPP	SQQ-SL	RAVA : 529
Q647H3_	:	KFFVPP	SQQ-SL	GAVA : 537
Q9SQH7_	:	KFFVPP	FQQ-SP	RAVA : 530
A1DZF0_	:	KFFVPP	FQQ-SP	RAVA : 529
O82580_	:	KFFVPP	SQQ-SP	RAVA : 507
Q5I6T2_	:	KFFVPP	SQQ-SP	RAVA : 531
B5TYU1_	:	KFFVPP	SQQ-SP	RAVA : 530

Figure S-3 B. Alignment of one abundant Ara h 3 isoform Q647H4_ARAHY with detected peptides from this study.

The signal sequence is underlined in italics, and the main sequence in bold and shaded where peptides were identified. Roasted peptides identified by DIA-IM-MS are aligned with the sequence, the most abundant peptides are underlined, raw peptides are in italics with the site of hydroxylation bold (**P**) and peptides found in both raw and roasted are in bold. The N-terminal sequences identified in purified Ara h 3 are marked (*)⁷.

SIGNAL SEQUENCE *MGKLLALSVCFCFLVLGA*

MATURE ACIDIC SUBUNIT

SSISFRQQPEENACQFQRLNAQRPDNRRIESEGYYIETWNPNNQEFECAGVALSRLVLRNALRRPFYSNAPQEIFIQQGR

QQPEENACQFQR

IESEGYYIETWNPNNQEFECAGVALSR

RPFYSNAPQEIFIQQGR

IESEGYYIETWNPNNQEFECAGVALSR

PFYSNAPQEIFIQQGR

LNAQRPDNR

NALRRPFYSNAPQEIFIQQGR

PDNR

LVLRR

LNAQR

LVLRR

LNAQRDNR

LNAQRPDNRRIESEGYYIETWNPNNQEFECAGVALSR

GYFGLIFPGCPSTYEPAQQGRRHQSQRPPRRFQGQDQSQQQQDSHQQVHRFDEGLIAVPTGVAFWMYNDHDTDVAVSLTDTNNNDN

GYFGLIFPGCPSTYEPAQQGR

RFQGQDQSQQQQDSHQQ

FDEGLIAVPTGVAFWMYNDHDTDVAVSLTDTNNNDN

GYFGLIFPGCPSTYEPAQQGR

FQGQDQSQQQQDSHQQ

GYFGLIFPGCPSTYEPAQQGRR

RFQGQDQSQQQQDSHQQVHR

RHQSQRPPR

HQSQRPPR

HQSQRPPR

HQSQRPPR

HQSQRPPR

QLDQFPRRFNLAGNHEQEFLRYQQSRRRSLPYSPYSPQTQPKQEDREFSPRGQHGRRERAGQEENEGGNIFSGFTPEFLAQAFQVDDR

QLDQFPRR

SLPYSPYSPQTQPK

GQHGR

AGQEENEGGNIFSGFTPEFLAQAFQVDDR

RFNLAGNHEQEFLR

SLPYSPYSPQTQPK

EFSPR

ERAGQEENEGGNIFSGFTPEFLAQAFQVDDR

FNLAGNHEQEFLR

PYSPYSPQTQPK

EFSPRGQHGR

FNLAGNHEQEFL

RSLPYSPYSPQTQPK

EFSPRGQHGR

YQQQSR RSLPYSPYSPQTQPK

YQQQSRR

QEDREFSPR

QEDREFSPR

QILQNLGENESDEQGAIVTVRGGLRILSPDRKRRQQYERPDEEEYDEDEYEDDEERQHDRRRRGRGSRGSGN
GENESDEQGAIVTVR **QQYERPDEEEYDEDEYEDDEER** **GSGN**
QILQNL GGLR *RQQYERPDEEEYDEDEYEDDEER* *GSGN*
GGLRILSPDRK *GSGN*
*QILQNL*GENESDEQGAIVTVR **ILSPDRK**
ILSPDR
ILSPDRK
ILSPDRKR

MATURE BASIC SUBUNIT SEQUENCE

GIEETICTASFKKNIGRNRSPDIYNPQAGSLKTANELNLLILRWLGLSAEYGNLYRNALFVPHYNTNAHSIIYALRGRAHVQVVD**SNGDR**
GIEETICTASFK **NRSPDIYNPQAGSLK** **WLGLSAEYGNLYR** *GRAHVQVVD***SNGDR**
GIEETICTASFKK **SPDIYNPQAGSLK** **AHVQVVD****SNGDR**
GIEETICTASFKKNIGR **TANELNLLILR** *NALFVPHYNTNAHSIIYALR* *AHVQVVD***SNGDR**
NIGRNRSPDIYNPQAGSLK *NALFVPHYNTNAHSIIYALRGR*
NRSPDIYNPQAGSLK
SPDIYNPQAGSLK
SPDIYNPQAGSLKTANELNLLILR

VFDEELQEGHVLVVPQNFAVAGKSQSENFHEYVAFKTD**SRPSIANLAGENSFIDNLPEEVVANSYGLPREQARQLKNNNP****FKFFVPPSEQSLRAVA**
VFDEELQEGHVLVVPQNFAVAGK **FFVPPSEQSLR**
VFDEELQEGHVLVVPQNFAVAGK *FFVPPSEQSLR*
SQSENFHEYVAFK *FFVPPSEQSLR*
*SQSENFHEYVAFKTD***SRPSIANLAGENSFIDNLPEEVVANSYGLPR** *QLKNNNP***FK**
VFDEELQEGHVLVVPQNFAVAGK *TDSRPSIANLAGENSFIDNLPEEVVANSYGLPREQAR* **NNNP****FK**
VFDEELQEGHVLVVPQNFAVAGK *TDSRPSIANLAGENSFIDNLPEEVVANSYGLPR*

Figure S-3 C. Alignment of a second abundant Ara h 3 isoform Q0GM57 with detected peptides from this study.

The signal sequence is underlined in italics, and the main sequence in bold and shaded where peptides were identified. Roasted peptides identified by DIA-IM-MS are aligned with the sequence, the most abundant peptides are underlined, raw peptides are in italics with the site of hydroxylation bold (**P**) and peptides found in both raw and roasted are in bold. The N-terminal sequences identified in purified Ara h 3 are marked (*)⁷.

SIGNAL SEQUENCE

MAKLLALSLCFCVLVLGA

MATURE ACIDIC SUBUNIT SEQUENCE

SSVTFRQGGEENECQFQRLNAQRPDNRRIESEGgyIETWNPNNQEFQCAGVALSRTVLRRNALRRPFYSNAPLEIYVQQGSGYF
QGGEENECQFQR IESEGgyIETWNPNNQEFQCAGVALSR RPFYSNAPLEIYVQQGSGYF...
LNAQRPDNR TVLRR RPFYSNAPLEIYVQQGSGYF ...
 PDNR
 QGGEENECQFQR
 LNAQR
 LNAQRPDNRRIESEGgyIETWNPNNQEFQCAGVALSR

GLIFPGCPSTYEPAQEGRRYQSQKPSRRFQVGQDDPSQQQQDSHQKVHRFDEGDLIAVPTGVAFWMYNDEDTDVVTVTLSDT
 GLIFPGCPSTYEPAQEGR FQVGQDDPSQQQQDSHQK FDEGDLIAVPTGVAFWMYNDEDTDVVTVTLSDT
 GLIFPGCPSTYEPAQEGR YQSQKPSRR FDEGDLIAVPTGVAFWMYNDEDTDVVTVTLSDT
 YQSQKPSR
 YQSQKPSR
 RYQSQKPSRR
 RYQSQKPSR

SSIHNQLDQFPRRFLYLAGNQEQEFLRYQQQQGSRPHYRQISPRVRGDEQENEGSNIFSGFAQEFLQHAFQVDRQTVENLRGENER
 SSIHNQLDQFPRR YQQQQGSRPHYR QTVENLRGENER
 SSIHNQLDQFPR FYLAGNQEQEFLR QTVENLR
 YQQQQGSRPHYRQISPR GENER
 YQQQQGSRPHYRQISPR GENER
 QTVENLRGENER

EEQGAIVTVKGLRILSPDEEDESSRSPPSRREEFDEDRSRPQQRGKYDENRRGYKN
EEQGAIVTVK ILSPDEEDESSR REEFDEDR N
GGLR EEFDEDR N
EEQGAIVTVKGLR GYKN
RGYKN

EEQGAIVTVK

GGLRILSPDEEDESSR
ILSPDEEDESSR
ILSPDEEDESSRSPPSRR
SPPSRR
SPPSRREEFDEDR
EEFDEDRSRPQQR

MATURE BASIC SUBUNIT SEQUENCE

GIEETICSASVKKNLGRSSNPDIYNPQAGSLRSVNELDLPILGWLGLSAQHGTIYRNAMFVPHYTLNAHTIVVALNGR
...GIEETICSASVK NLGRSSNPDIYNPQAGSLR NAMFVPHYTLNAHTIVVALNGR
...GIEETICSASVK NLGRSSNPDIYNPQAGSLR
...GIEETICSASVKK SSNPDIYNPQAGSLR
...GIEETICSASVK SVNELDLPILGWLGLSAQHGTIYR
SVNELDLPILGWLGLSAQHGTIYR

AHVQVVDSSNGNRVYDEELQEGHVLVVPQNFAVAAKAQSENYEYLAFKTDSRPSIANLAGENSIIDNLPEEVVANSYRLPREQARQLKNNNPFKFFVPPFDHQSMREVA
VYDEELQEGHVLVVPQNFAVAAK TDSRPSIANLAGENSIIDNLPEEVVANSYR FFVPPFDHQSMR
AHVQVVDSSNGNR TDSRPSIANLAGENSIIDNLPEEVVANSYRLPR FFVPPFDHQSMR
TDSRPSIANLAGENSIIDNLPEEVVANSYR NNNPFK
TDSRPSIANLAGENSIIDNLPEEVVANSYR QLKNNNPFK
AQSENYEYLAFK NNNPFKFFVPPFDHQSMR
LPREQARQLK

Figure S-4 A. Alignment of Ara h 2 sequences used in this study from the compressed UniProt accession Q6PSU2 (* denotes every 10th residue).

		*	20	*	40	*	60	*	80	*	100	*		
Q6PSU2	:	MAKLTILVALALFLLAAHASARQQWELQGDRRCQSQLERANLRPCEQHLMQKIQRDEDSYGRDPYSPSQDFYSPSQDPDRRDEYSPSPYDRRGAGSSQHQERCCNELNEF											:	110
Q6PSU2-3	:	MAKLTILVALALFLLAAHASARQQWELQGDRRCQSQLERANLRPCEQHLMQKIQRDEDSYGRDPYSPSQDFYSPSQDPDRRDEYSPSPYDRRGAGSSQHQERCCNELNEF											:	110
Q6PSU2-2	:	MAKLTILVALALFLLAAHASARQQWELQGDRRCQSQLERANLRPCEQHLMQKIQRDEDSYGRDPYSPSQDF-----YSPSPYDRRGAGSSQHQERCCNELNEF											:	98
Q6PSU2-4	:	MAKLTILVALALFLLAAHASARQQWELQGDRRCQSQLERANLRPCEQHLMQKIQRDEDSYGRDPYSPSQDF-----YSPSPYDRRGAGSSQHQERCCNELNEF											:	98

		120	*	140	*	160	*		
Q6PSU2	:	ENNQRCMCEALQQIMENQSDRLQGRQQEQQFKRELRLNLPQQCGLRAPQRCDEVESGGRDRY						:	172
Q6PSU2-3	:	ENNQRCMCEALQQIMENQSDRLQGRQQEQQFKRELRLNLPQQCGLRAPQRCDEVESGGRD~~						:	170
Q6PSU2-2	:	ENNQRCMCEALQQIMENQSDRLQGRQQEQQFKRELRLNLPQQCGLRAPQRCDEVESGGRDRY						:	160
Q6PSU2-4	:	ENNQRCMCEALQQIMENQSDRLQGRQQEQQFKRELRLNLPQQCGLRAPQRCDEVESGGRD~~						:	158

Figure S-4 B. Alignment of an Ara h 2 isoform Q6PSU2 with detected peptides from this study.

The signal sequence is underlined in italics, and the main sequence in bold and shaded where peptides were identified. Roasted peptides identified by DIA-IM-MS are aligned with the sequence, the most abundant peptides are underlined, raw peptides are in italics with the site of hydroxylation bold (**P**) and peptides found in both raw and roasted are in bold. The N-terminal sequences identified in purified Ara h 2 are marked (*)⁷. One peptide (⁶³D-R⁸²) spanned the insert region found in two of the Q6PSU2 isoforms whilst a second peptide was observed at the C-terminus with an arginine residue (¹⁶²C-R^{171/159}).

SIGNAL SEQUENCE MAKLILVALALFLLAAHASA

MATURE SEQUENCE

RQQWELQGDRRCQSLERANLRPCEQHLMQKIQRDEDSYGRDPYSPSQDPYSPSQDPDRRDYSPSPYDRRGAGSSQHQR

QQWELQGDRR **ANLRPCEQHLMQK** **DPYSPSQDPYSPSQDPDRR** **RGAGSSQHQR**

QQWELQGR **CQSLER** **IQRDEDSYGR** **DPYSPSPYDR** **GAGSSQHQR**

RCQSLER **PCEQHLMQK** *DPYSPSQDPYSPSQDPDRRDYSPSPYDR*

ANLRPCEQHLMQK *DPYSPSQDPYSPSQDPDRR*

RDYSPSPYDRR

RDYSPSPYDR

RDYSPSPYDR

RDYSPSPYDRR

DPYSPSPYDRR

DPYSPSPYDR

IQRDEDSYGRDPYSPSQDPYSPSQDPDRR

IQRDEDSYGRDPYSPSQDPYSPSQDPDR

DEDSYGRDPYSPSQDPYSPSQDPDRR

DEDSYGR

DEDSYGRDPYSPSQDPYSPSQDPDR

DEDSYGRDPYSPSQDPYSPSQDPDRR

CCNELNEFENNQR **CMCEALQQIMENQSDRLQGRQQEQQFKRELRNLPQQCGLRAPQRCDLEVESGGRDRY**

CCNELNEFENNQR **LQGRQQEQQFK** **ELRNLPQQCGLR**

CCNELNEFENNQR *LQGRQQEQQFKR* **NLPQQCGLRAPQR**

QQIMENQSDR **QQEQQFK** **NLPQQCGLR** **CDLEVESGGR**

QQEQQFK *NLPQQCGLR* **CDLEVESGGRDR**

CMCEALQQIMENQSDR *QQEQQFKRELR* **APQRCDLEVESGGR**

CMCEALQQIMENQSDRLQGR

Figure S-5 A. Alignment of Ara h 6 sequences used in this study (* denotes every 10th residue).

		*	20	*	40	*	60	*	80		
Q647G9_	:	MAKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLKPCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCC								:	80
A1DZE9_	:	MAKSTILVALLALVLVAHASAMRRERGRQGDSSSCERQVDRVNLKPCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCC								:	80

		*	100	*	120	*	140		
Q647G9_	:	DELNEMENTQRCMGEALQQIMENQCDRLQDRQMVFQFKREL MNLPQQCNFRAPQRCDLDVSGGRC						:	145
A1DZE9_	:	DELNEMENTQRCMGEALQQIMENQCDKLQDREMVEQFKRKLMDLAQQCNFRAPERCDLDVNGGRC						:	145

Figure S-5 B. Alignment of an Ara h 6 isoform Q647G9 with detected peptides from this study.

The signal sequence is underlined in italics and the main sequence in bold and shaded where peptides were identified. Roasted peptides identified by DIA-IM-MS are aligned with the sequence, the most abundant peptides are underlined, raw peptides are in italics with the site of hydroxylation bold (**P**) and peptides found in both raw and roasted are in bold. The N-terminal sequences identified in purified Ara h 6 are marked (*)⁷. Three peptides spanned regions unique to the Q647G9 and A5Z1R0 isoforms (⁹⁴C-R¹⁰⁹, ¹²²E-R¹³¹ and ¹³⁸C-R¹⁴⁴).

SIGNAL SEQUENCE MAKSTILVALLALVLVAHASA

MATURE SEQUENCE

MRRERGRQGDSSSCERQVDRVNLKPCEQHIMQRIMGEQEYDSYDIRSTRSSDQQQRCCDELNEMENTQRCMCEALQQIMENQCDRLQDR

IMGEQEYDSYDIR

CMCEALQQIMENQCDR

VNLKPCEQHIMQR

CMCEALQQIMENQCDRLQDR

PCEQHIMQR

CCDELNEMENTQR

QVDRVNLKPCEQHIMQR

SSDQQQRCCDELNEMENTQR

VNLKPCEQHIMQR

SSDQQQR

LQDR

GRQGDSSSCER

PCEQHIMQR

STRSSDQQQR

GRQGDSSSCERQVDR

QGDSSSCER

QGDSSSCERQVDR

QMVQQFKRELMNLPQQCNFRAPQRCDLDVSGGR

RELMNLPQQCNFR

CDLDVSGGR

ELMNLPQQCNFR

QMVQQFKR

APQRCDLDVSGGR

QMVQQFK

APQRCDLDVSGGR

QMVQQFK

QMVQQFKRELMNLPQQCNFR

QMVQQFKRELMNLPQQCNFR

RELMNLPQQCNFR

ELMNLPQQCNFR

MVQQFKR

MVQQFK

Table S-3. Peanut allergens profiled in three samples of raw peanuts and two batches roasted peanut flour using DIA-IM-MS.

^aRelative molecular mass ⁷; ^bThe mean average pairwise identities of the allergen sequences in each protein, ^cUniProt accession number (redundant sequences in the database are noted); ^dConfidence scores are calculated from the combined scores of peptides contributing to peptide ID; ^emean and standard deviation of the % sequence coverage observed in all replicate analyses of the given peanut sample. The following sequences were determined to be redundant (i.e. identical): ^fP43238 and N1NG13; ^gP43237 and B3IXL2; ^hQ647G9 and A5Z1R0. ND, not detected. Identification of isoforms for which only a single mass event was observed were confirmed by visual inspection of the spectra. No peptides were identified corresponding to the Profilin family allergen Ara h 5, accessions Q9SQI9, D3K177 and L7QH52. EIC for protein isoforms which were identified using a single peptide ID are shown in Figure S-6.

Allergen name	Subunit Mr (kDa) ^a	Average identity (%) ^b	UniProt accession ^c	Raw Peanut					Roasted Peanut			
				Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e			Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e	
						Raw 1	Raw 2	Raw 3			Batch 1	Batch 2
Cupin barrel family												
Ara h 1	63.5	95.22	P43238 ^f	303 (0)	1305.5	76.3 +/- 1.9	78.0 +/- 0.3	76.1 +/- 0.6	236 (0)	1314.3	71.2 +/- 2.9	71.7 +/- 0.3
			P43237 ^g	285 (0)	1427.6	77.6 +/- 1.2	77.5 +/-0.6	77.8 +/- 0.7	213 (1)	1258.4	70.9 +/- 3.6	72.1 +/- 2.4
			E5G076	229 (3)	1058.4	63.6 +/- 3.1	63.6 +/- 3.3	60.4 +/- 1.3	182 (6)	1084.1	57.4 +/-5.7	57.9 +/- 0.1
			Q6PSU3	281 (0)	1396.9	78.1 +/- 1.3	78.8 +/- 0.5	78.5 +/- 0.2	206 (0)	1223.3	72.1 +/- 2.5	72.8 +/- 2.7

Allergen name	Subunit Mr (kDa) ^a	Average identity (%) ^b	UniProt accession ^c	Raw Peanut					Roasted Peanut			
				Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e			Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e	
						Raw 1	Raw 2	Raw 3			Batch 1	Batch 2
Ara h 3	40 (acidic), 20 (basic)	82.76	Q5I6T2	241 (16)	936.7	82.3 +/- 2.8	82.2 +/- 1.7	81.6 +/- 2.2	186 (18)	802.2	67.5 +/- 4.0	66.6 +/- 4.7
			B5TYU1	249 (4)	980.6	85.5 +/- 4.2	85.4 +/- 3.3	84.7 +/- 4.0	174 (2)	815.4	72.7 +/- 6.0	72.5 +/- 6.4
			Q9FZ11	236 (8)	909.7	76.4 +/- 1.5	74.8 +/- 0.5	75.9 +/- 1.6	184 (11)	801.5	68.4 +/- 3.2	66.6 +/- 6.1
			Q647H4	240 (1)	840.3	83.8 +/- 0.9	82.8 +/- 3.2	84.9 +/- 2.6	168 (1)	756.4	73.6 +/- 2.8	71.1 +/- 6.8
			Q647H3	226 (8)	893.2	80.2 +/- 2.4	77.5 +/- 1.8	79.2 +/- 0.5	172 (1)	737.9	69.6 +/- 5.7	69.4 +/- 7.3
			Q8LKN1	239 (0)	859.6	77.8 +/- 1.1	78.7 +/- 0.8	78.8 +/- 0.3	161 (0)	730.3	70.0 +/- 2.7	66.5 +/- 6.4
			Q6T2T4	229 (1)	829.1	75.8 +/- 1.1	76.8 +/- 0.8	76.7 +/- 0.3	156(0)	712.3	68.2 +/- 2.7	64.6 +/- 6.4
			Q9SQH7	158(10)	553.2	59.5 +/- 4.5	59.8 +/- 4.6	59.4 +/- 5.5	125 (9)	458.2	46.5 +/- 3.8	50.4 +/- 7.8
			Q0GM57	118 (1)	550.0	70.6 +/- 6.3	53.4 +/- 2.2	71.1 +/- 2.2	128 (2)	615.4	64.0 +/- 5.5	62.9 +/- 7.0
			E5G077	119 (10)	540.4	61.2 +/- 7.1	46.0 +/- 1.6	62.2 +/- 3.9	115 (0)	524.1	56.7 +/- 5.5	55.2 +/- 6.3
			Q647H2	40 (34)	299.3	45.3 +/- 5.0	48.5 +/- 1.8	50.5 +/- 2.9	35 (25)	216.9	36.3 +/- 6.0	38.5 +/- 2.7
			A1DZF0	258 (63)	1032.3	86.1 +/- 4.2	85.9 +/- 3.5	85.1 +/- 4.1	170 (31)	770.3	72.5 +/- 6.0	70.7 +/- 9.1

Allergen name	Subunit Mr (kDa) ^a	Average identity (%) ^b	UniProt accession ^c	Raw Peanut					Roasted Peanut			
				Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e			Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e	
						Raw 1	Raw 2	Raw 3			Batch 1	Batch 2
Prolamin family												
Ara h 2	17	92.96	Q6PSU2	127 (0)	506.5	85.0 +/- 1.3	85.8 +/- 1.7	86.2 +/- 0.3	131 (0)	754.2	80.9 +/- 0.1	81.9 +/- 1.5
			Q6PSU2-2	79 (0)	341.0	71.3 +/- 1.9	75.2 +/- 5.6	73.5 +/-1.4	88 (0)	511.8	68.2 +/- 0.1	69.3 +/- 1.6
			Q6PSU2-3	120 (0)	488.0	84.6 +/- 1.3	85.7 +/- 1.8	85.6 +/- 0.7	130 (0)	729.4	80.7 +/- 0.2	81.8 +/- 1.6
			Q6PSU2-4	72 (0)	322.6	70.7 +/- 2.0	74.8 +/- 5.6	73.3 +/- 1.1	87 (0)	487.0	67.8 +/- 0.2	69.0 +/- 1.8
Ara h 6	16	93.10	Q647G9 ^h	45 (0)	252.4	71.6 +/- 3.4	75.5 +/- 0.4	72.5 +/- 1.3	60 (0)	445.6	68.6 +/- 2.4	66.6 +/- 2.5
			A1DZE9	34 (6)	230.4	46.4 +/- 9.3	54.7 +/- 1.7	62.5 +/- 8.1	39 (11)	313.1	40.2 +/- 8.6	42.2 +/- 1.4
Ara h 7	14	83.17	Q9SQH1	17(0)	65.4	23.1 +/- 0.0	38.8 +/- 0.0	23.1 +/- 0.0	ND	ND	ND	ND
			Q647G8	50 (21)	268.5	76.5 +/- 1.2	74.2 +/- 2.7	77.4 +/- 1.5	24 (7)	184.1	56.4 +/- 10.3	62.9 +/- 6.6
			B4XID4	69 (32)	335.7	79.7 +/- 2.8	78.4 +/- 4.0	81.0 +/- 2.2	41 (21)	321.3	56.2 +/-10.5	62.5 +/- 8.5
Ara h 9	10		B6CEX8	10 (4)	80.4	46.9 +/- 1.9	49.8 +/- 4.0	49.6 +/- 2.9	9 (5)	70.4	55.7 +/- 3.4	56.3 +/- 2.4
			B6CG41	12 (7)	106.4	94.2 +/- 1.0	95.3 +/- 3.1	95.3 +/- 1.9	11 (6)	82.2	81.3 +/- 5.2	83.1 +/- 4.9

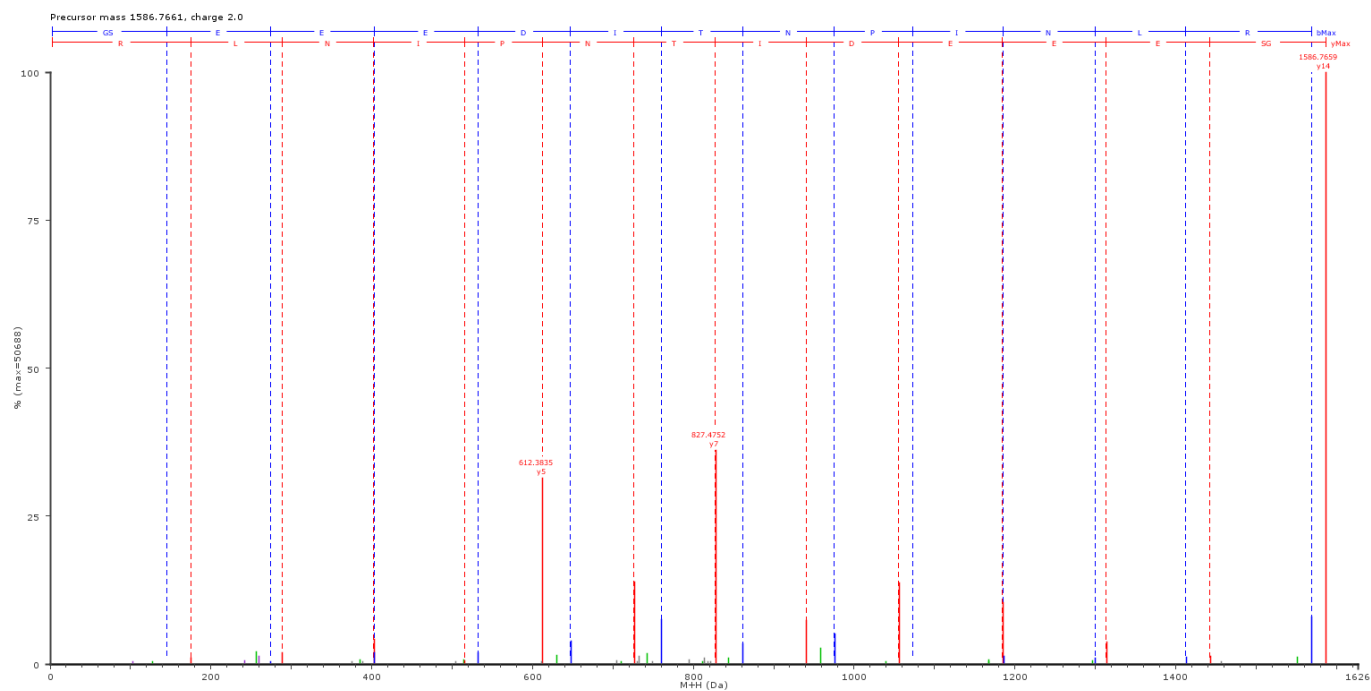
Allergen name	Subunit Mr (kDa) ^a	Average identity (%) ^b	UniProt accession ^c	Raw Peanut					Roasted Peanut			
				Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e			Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e	
						Raw 1	Raw 2	Raw 3			Batch 1	Batch 2
Bet v 1 family												
Ara h 8	17	80.08	Q6VT83	21 (1)	127.3	51.1 +/- 5.8	49.4 +/- 2.0	45.8 +/- 0.7	3 (0)	11.3	31.2 +/- 0.0	ND
			B2ZGS2	15 (1)	86.8	34.5 +/- 5.6	33.8 +/- 4.5	31.1 +/- 4.2	ND	ND	ND	ND
			B1PYZ4	30 (5)	197.0	64.0 +/- 4.9	64.7 +/- 4.5	62.3 +/- 1.4	5 (1)	19.7	31.2 +/- 0.0	ND
			Q0PKR4	24 (10)	161.4	ND	49.8 +/- 3.2	34.4 +/- 0.0	ND	ND	ND	ND
			Q5XXQ4	12 (3)	89.4	69.9 +/- 4.6	34.4 +/- 5.2	34.3 +/- 2.2	14 (3)	62.0	29.1 +/- 9.9	27.6 +/- 17
			B0YIU5	15 (5)	106.7	53.1 +/- 5.4	89.5 +/- 5.4	49.6 +/- 0.7	11 (3)	57.3	40.0 +/- 10.9	41.6 +/- 11.7

Allergen name	Subunit Mr (kDa) ^a	Average identity (%) ^b	UniProt accession ^c	Raw Peanut					Roasted Peanut			
				Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e			Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e	
						Raw 1	Raw 2	Raw 3			Batch 1	Batch 2
Oleosin family												
Ara h 10 and 11	15.5	73.70	Q647G5	35 (18)	213.8	42.3 +/- 0.8	43.0 +/- 0.2	42.5 +/- 0.5	16 (9)	127.2	35.7 +/- 1.5	35.5 +/- 1.7
			Q647G4	9 (0)	71.9	23.9 +/- 0.3	24.4 +/- 0.7	24.5 +/- 0.5	7 (0)	76.1	23.5 +/- 0.3	23.6 +/- 0.4
			Q9AXI0	17 (3)	118.2	41.5 +/- 1.9	41.5 +/- 1.6	41.6 +/- 0.7	9 (4)	50.6	9.5 +/- 5.4	20.5 +/- 0.0
			Q9AXI1	14 (0)	101.9	35.2 +/- 0.1	36.4 +/- 1.0	36.2 +/- 1.0	6 (0)	40.1	6.3 +/- 0.0	21.3 +/- 0.0
			Q6J1J8	19 (0)	130.0	42.7 +/- 1.2	42.7 +/- 0.5	42.8 +/- 1.9	10 (3)	59.5	6.3 +/- 0.0	32.4 +/- 0.0
			Q45W86	19 (7)	140.0	36.3 +/- 1.3	36.0 +/- 1.0	34.9 +/- 2.1	7 (2)	39.8	16.1 +/- 0.0	ND
			Q647G3	11 (8)	79.0	41.8 +/- 0.7	44.3 +/- 1.0	42.5 +/- 2.2	3 (2)	12.2	ND	21.7 +/- 0.0
			Q45W87	16 (5)	103.1	35.9 +/- 1.6	36.0 +/- 0.0	37.0 +/-2.3	7 (2)	41.2	16.1 +/- 0.0	20.4 +/- 0.0

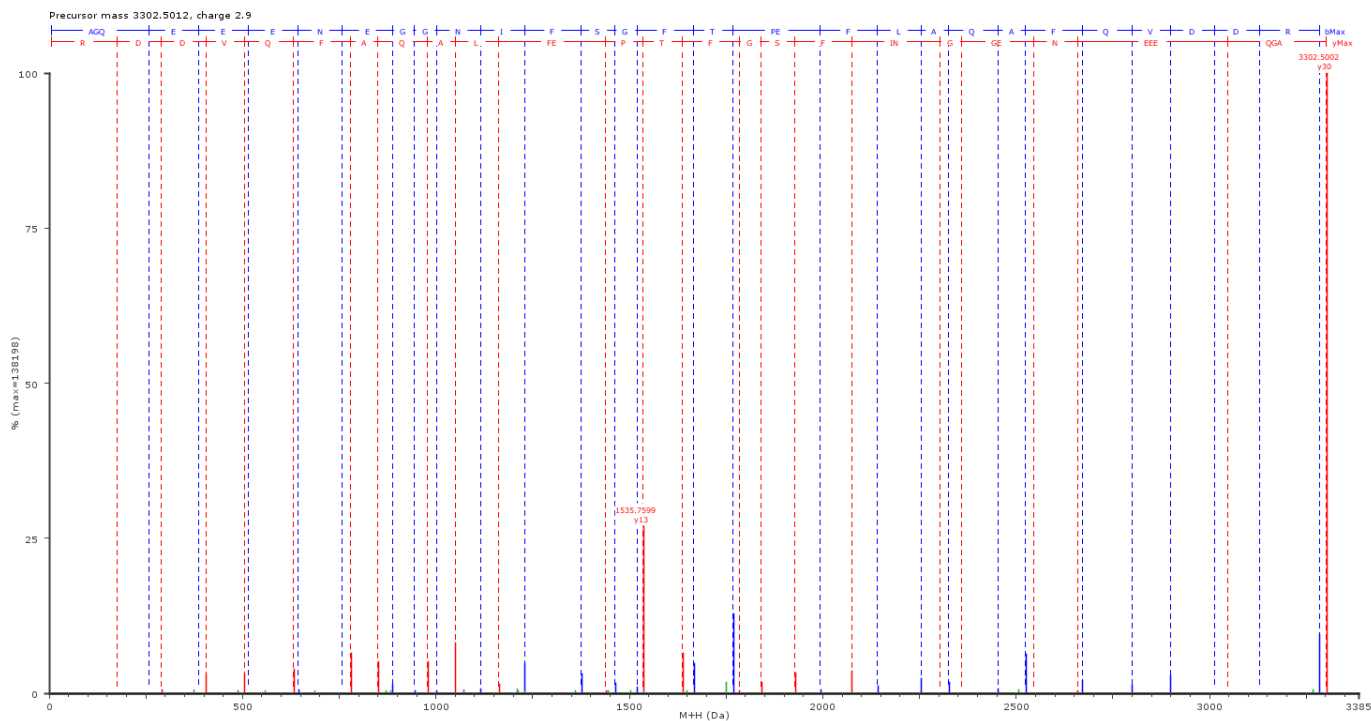
Allergen name	Subunit Mr (kDa) ^a	Average identity (%) ^b	UniProt accession ^c	Raw Peanut					Roasted Peanut			
				Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e			Number of peptides (unique)	Confidence Score ^d	Sequence coverage (%) ^e	
						Raw 1	Raw 2	Raw 3			Batch 1	Batch 2
Profilin family												
Ara h 5	14.0	94.91	Q9SQI9	ND	ND	ND	ND	ND	ND	ND	ND	ND
			D3K177	ND	ND	ND	ND	ND	ND	ND	ND	ND
			L7QH52	ND	ND	ND	ND	ND	ND	ND	ND	ND

Figure S-6 Extracted ion chromatogram for protein isoforms which were identified using a single peptide ID

(A) Ara h 1 P43237 (roasted peanut flour)



(B) Ara h 3 Q6T2T4 (raw)



(C) Ara h 8 Q6VT83(raw)



(D) Ara h 8 B2ZGS2 (raw)

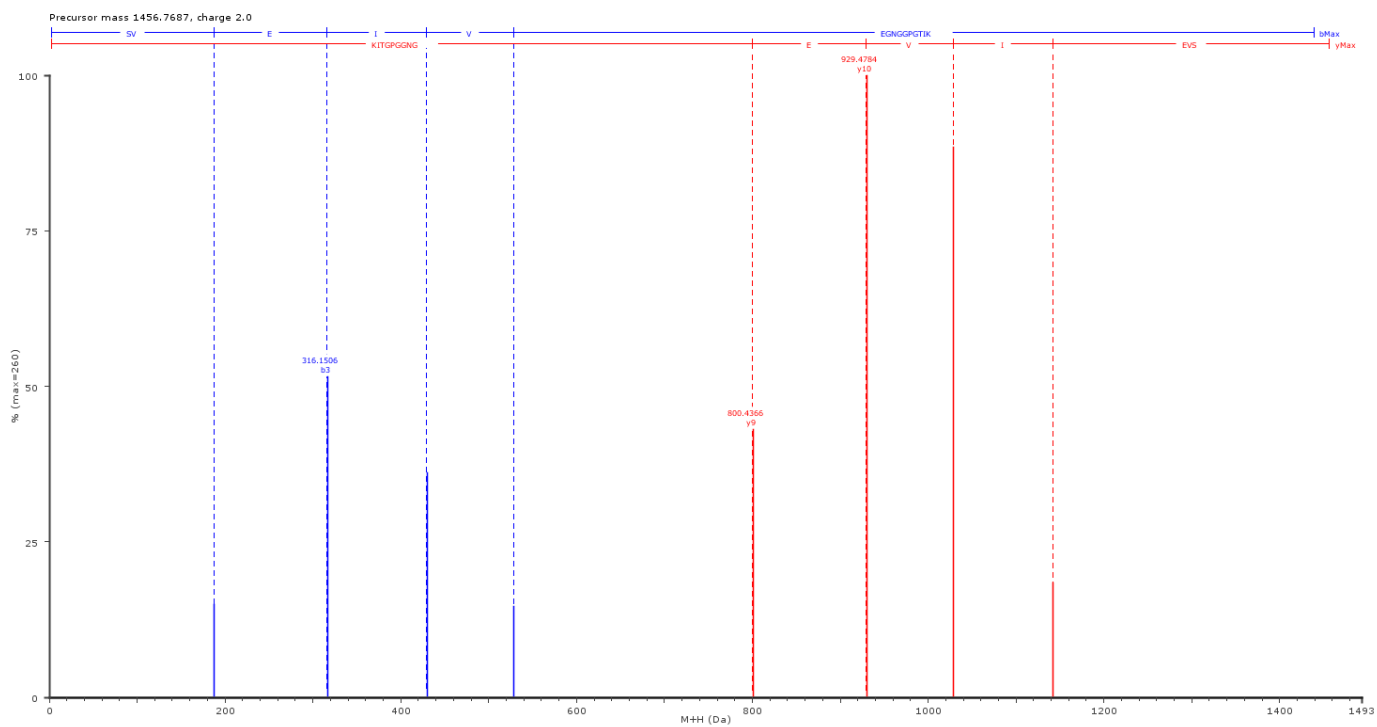


Table S-4. Peanut proteins (other than allergens) identified in three samples of raw peanuts and two batches roasted peanut flour using DIA-IM-MS.

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
A0A023IUM3	Histone deacetylase OS=Arachis hypogaea GN=HDA1 PE=2 SV=1	-	-	5	3
A0A023IUN7	Histone deacetylase OS=Arachis hypogaea PE=2 SV=1	7	4	-	-
A0A061	Lactoylglutathione lyase OS=Arachis hypogaea GN=GLYI PE=2 SV=2	-	-	8	5
A0A075BIU4	60S ribosomal protein L4/L1 OS=Arachis hypogaea PE=2 SV=1	31	1	-	-
A0A075M981	Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit (Fragment) OS=Arachis hypogaea GN=rbcL PE=3 SV=1	48	0	30	0
A0A088BAX2	Diacylglycerol acyltransferase OS=Arachis hypogaea GN=DGAT PE=2 SV=1	-	-	13	7
A0A089ZXL7	Peanut agglutinin variant OS=Arachis hypogaea PE=2 SV=1	26	3	17	2
A0A0A1ESG0	Actin (Fragment) OS=Arachis hypogaea PE=3 SV=1	21	0	6	0
A0A0A1ESX4	Actin (Fragment) OS=Arachis hypogaea PE=4 SV=1	6	1	-	-
A0A0A1EUV7	Actin (Fragment) OS=Arachis hypogaea PE=3 SV=1	31	11	2	0
A0A0A1EXN0	Actin (Fragment) OS=Arachis hypogaea PE=3 SV=1	24	1	8	1
A0A0A6ZDN9	Peptidyl-prolyl cis-trans isomerase OS=Arachis hypogaea PE=2 SV=1	18	4	9	1
A0A0A6ZDP1	Glyceraldehyde-3-phosphate dehydrogenase C2 OS=Arachis hypogaea PE=2 SV=1	25	4	13	1
A0A0A6ZDR7	Ribosomal protein L15 OS=Arachis hypogaea PE=2 SV=1	5	4	-	-
A0A0A6ZDR9	Serine hydroxymethyltransferase OS=Arachis hypogaea PE=2 SV=1	37	3	-	-
A0A0A6ZDS3	60S ribosomal protein L4/L1 OS=Arachis hypogaea PE=4 SV=1	27	0	-	-

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
A0A0A6ZDS5	Nucleoside diphosphate kinase OS=Arachis hypogaea PE=2 SV=1	16	3	-	-
A0A0A6ZDS7	60s ribosomal protein L1 OS=Arachis hypogaea PE=2 SV=1	29	0	-	-
A0A0A6ZDT0	Glyceraldehyde-3-phosphate dehydrogenase C2 OS=Arachis hypogaea PE=3 SV=1	25	3	13	1
A0A0A6ZDT3	Serine hydroxymethyltransferase OS=Arachis hypogaea PE=3 SV=1	35	1	-	-
A0A0A6ZDT5	60S ribosomal protein L1 OS=Arachis hypogaea PE=3 SV=1	34	6	-	-
A0A0A6ZDX5	40s ribosomal protein S30 OS=Arachis hypogaea PE=2 SV=1	1	0	-	-
A0A0A6ZDX8	40S ribosomal protein S3 OS=Arachis hypogaea PE=2 SV=1	8	5	-	-
A0A0A6ZDY0	40S ribosomal protein S30 OS=Arachis hypogaea PE=4 SV=1	1	0	-	-
A0A0A6ZDY5	Nucleoside diphosphate kinase OS=Arachis hypogaea PE=3 SV=1	13	0	-	-
A0A0A7E0N9	Uncharacterized protein OS=Arachis hypogaea PE=2 SV=1	-	-	4	3
A0A0C4KCU2	Ribulose biphosphate carboxylase large chain OS=Arachis hypogaea GN=rbcL PE=3 SV=1	54	2	34	3
A0A0C4KCX1	Cytochrome f OS=Arachis hypogaea GN=petA PE=3 SV=1	-	-	5	3
A0A0C4KCY1	50S ribosomal protein L20, chloroplastic OS=Arachis hypogaea GN=rpl20 PE=3 SV=1	-	-	4	1
A0A0C4KDA8	ATP synthase epsilon chain, chloroplastic OS=Arachis hypogaea GN=atpE PE=3 SV=1	-	-	3	3
A0A0C4KDE7	Photosystem II reaction center protein L OS=Arachis hypogaea GN=psbL PE=3 SV=1	1	1	-	-
A0A0C4KDH4	30S ribosomal protein S8, chloroplastic OS=Arachis hypogaea GN=rrs8 PE=3 SV=1	1	1	-	-
A0A0C4KJ20	50S ribosomal protein L14, chloroplastic OS=Arachis hypogaea GN=rpl14 PE=3 SV=1	-	-	1	1

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
A0A0C4KJ56	Ycf1 protein OS=Arachis hypogaea GN=ycf1 PE=4 SV=1	68	41	-	-
A0A0C4KPC6	ATP synthase subunit beta, chloroplastic OS=Arachis hypogaea GN=atpB PE=3 SV=1	11	9	-	-
A0A0C4KPM0	Photosystem I iron-sulfur center OS=Arachis hypogaea GN=psaC PE=3 SV=1	-	-	2	2
A0A0C5K1J5	LEC1 protein OS=Arachis hypogaea GN=LEC1 PE=2 SV=1	-	-	6	0
A0A0D4CHG2	Lipoxygenase OS=Arachis hypogaea GN=LOX6_C05 PE=2 SV=1	51	2	-	-
A0A0D4CHN0	Lipoxygenase OS=Arachis hypogaea GN=LOX6_E10 PE=2 SV=1	48	0	-	-
A0A0D4CI32	Lipoxygenase OS=Arachis hypogaea GN=LOX7_G2 PE=2 SV=1	36	20	12	4
A0A0D4CI71	Lipoxygenase (Fragment) OS=Arachis hypogaea GN=LOX8_H07 PE=2 SV=1	59	29	-	-
A6XN43	Actin (Fragment) OS=Arachis hypogaea PE=2 SV=1	2	0	-	-
A7LB59	Steroleosin-B OS=Arachis hypogaea GN=STO-B PE=2 SV=1	38	31	32	28
A7LB60	Steroleosin A OS=Arachis hypogaea GN=STO-A PE=2 SV=1	30	26	17	12
A9P5P3	Acyl ACP-thioesterase OS=Arachis hypogaea GN=pFatB PE=2 SV=1	-	-	11	8
A9YSZ2	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (Fragment) OS=Arachis hypogaea GN=rbcL PE=3 SV=1	9	0	8	1
B0LX54	Resveratrol synthase OS=Arachis hypogaea PE=3 SV=1	8	0	-	-
B0ZZ79	Potyvirus VPg interacting protein OS=Arachis hypogaea GN=VPgB PE=2 SV=1	17	9	-	-
B2LR85	NBS-LRR type disease resistance protein GnRGC12 (Fragment) OS=Arachis hypogaea PE=4 SV=1	-	-	1	0
B4UW48	Putative uncharacterized protein (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	3	1

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
B4UW50	Chitinase-like protein (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	5	1
B4UW54	GroES-like protein (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	3	2
B4UW57	Putative dihydroflavonol reductase (Fragment) OS=Arachis hypogaea PE=2 SV=1	12	7	5	4
B4UW59	Double-stranded DNA-binding protein OS=Arachis hypogaea PE=2 SV=1	6	3	-	-
B4UW61	Electron transporter/thiol-disulfide exchange intermediate protein OS=Arachis hypogaea PE=2 SV=1	-	-	3	1
B4UW62	Embryo-abundant protein EMB (Fragment) OS=Arachis hypogaea PE=2 SV=1	4	2	-	-
B4UW64	Epoxide hydrolase (Fragment) OS=Arachis hypogaea PE=2 SV=1	10	6	-	-
B4UW68	Putative ferredoxin (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	2	2
B4UW70	Fiber annexin (Fragment) OS=Arachis hypogaea PE=2 SV=1	24	19	3	2
B4UW74	Fructokinase (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	4	3
B4UW79	Glutathione peroxidase (Fragment) OS=Arachis hypogaea PE=2 SV=1	19	14		
B4UW82	Glutathione S-transferase 3 (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	5	2
B4UW88	Heat shock protein 1 (Fragment) OS=Arachis hypogaea PE=2 SV=1	23	1	10	0
B4UW89	Heat shock protein 2 OS=Arachis hypogaea PE=2 SV=1	24	3	10	0
B4UW91	Putative heat shock protein 4 (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	9	8
B4UWA1	Lactoylglutathione lyase (Fragment) OS=Arachis hypogaea PE=2 SV=1	21	16	23	16
B4UWA4	Putative uncharacterized protein OS=Arachis hypogaea PE=2 SV=1	1	1		

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
B4UWA8	Hypoxanthine phosphoribosyltransferase OS=Arachis hypogaea PE=2 SV=1	-	-	7	4
B4UWA9	Putative inorganic pyrophosphatase (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	2	0
B4UWB2	Kunitz trypsin inhibitor 4 OS=Arachis hypogaea PE=2 SV=1	9	7	-	-
B4UWB9	Lipoxygenase 1 (Fragment) OS=Arachis hypogaea PE=2 SV=1	17	9	-	-
B4UWC0	Lipoxygenase 2 (Fragment) OS=Arachis hypogaea PE=2 SV=1	15	0	-	-
B4UWC2	Gonadotropin beta chain OS=Arachis hypogaea PE=2 SV=1	-	-	6	5
B4UWD2	Myo-inisitol oxygenase (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	9	5
B4UWD3	Putative mitochondrial ATP synthase (Fragment) OS=Arachis hypogaea PE=2 SV=1	3	2	-	-
B4UWD4	Putative mRNA export protein (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	5	1
B4UWD9	Threonine endopeptidase (Fragment) OS=Arachis hypogaea PE=2 SV=1	3	2	-	-
B4UWE4	Thioredoxin fold (Fragment) OS=Arachis hypogaea PE=2 SV=1	4	0	-	-
B4YA12	Calcium calmodulin-dependent protein kinase (Fragment) OS=Arachis hypogaea GN=CCaMK PE=2 SV=1	-	-	6	3
B5LNQ2	Pathogenesis-induced protein OS=Arachis hypogaea GN=PIP PE=2 SV=1	6	3	-	-
B5TKB7	Actin (Fragment) OS=Arachis hypogaea PE=2 SV=1	2	0	-	-
B6ECZ0	Acyl-[acyl-carrier-protein] desaturase OS=Arachis hypogaea PE=2 SV=1	-	-	1	0
C0L2U6	Putative RNA-directed DNA polymerase OS=Arachis hypogaea PE=4 SV=1	-	-	6	3
D3Y2R1	Acyl carrier protein OS=Arachis hypogaea GN=ACP4 PE=2 SV=1	5	4	-	-

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
D6C528	Leafy cotyledon 1-A OS=Arachis hypogaea GN=LEC1-A PE=2 SV=1	-	-	7	1
D6C529	LEC1 protein OS=Arachis hypogaea GN=LEC1-B PE=2 SV=1	-	-	3	1
D8KXY1	Acyl carrier protein OS=Arachis hypogaea GN=ACP1 PE=2 SV=1	10	7		
D8KXZ2	Beta-ketoacyl-ACP reductase 1-1 OS=Arachis hypogaea GN=KR1-1 PE=2 SV=1	-	-	3	0
D8KXZ3	Beta-ketoacyl-ACP reductase OS=Arachis hypogaea GN=KR1-2 PE=2 SV=1	-	-	3	0
D8KXZ4	Beta-ketoacyl-ACP reductase 1-3 OS=Arachis hypogaea GN=KR1-3 PE=2 SV=1	-	-	3	0
E0WN93	Cystein proteinase inhibitor OS=Arachis hypogaea GN=cystatin PE=4 SV=1	1	0	-	-
E2IQH6	Sodium/hydrogen exchanger OS=Arachis hypogaea GN=NHX1 PE=2 SV=1	3	1	5	1
E4W7U8	Putative resveratrol synthesis enzyme OS=Arachis hypogaea GN=resveratrol synthesis PE=3 SV=1	11	0	-	-
E5FHY0	Late embryogenesis abundant protein group 1 protein OS=Arachis hypogaea GN=LEA1-1 PE=2 SV=1	24	0	17	0
E5FHY1	Late embryogenesis abundant protein group 1 protein OS=Arachis hypogaea GN=LEA1-2 PE=2 SV=1	26	10	13	4
E5FHY2	Late embryogenesis abundant protein group 1 protein OS=Arachis hypogaea GN=LEA1-3 PE=2 SV=1	23	2	13	0
E5FHY3	Late embryogenesis abundant protein group 2 protein dehydrin OS=Arachis hypogaea GN=DHN1 PE=2 SV=1	21	15	18	12
E5FHY4	Late embryogenesis abundant protein group 3 protein OS=Arachis hypogaea GN=LEA3-1 PE=2 SV=1	40	35	19	15
E5FHY5	Late embryogenesis abundant protein group 3 protein OS=Arachis hypogaea GN=LEA3-2	9	3	-	-

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
	PE=2 SV=1				
E5FHY6	Late embryogenesis abundant protein group 3 protein (Fragment) OS=Arachis hypogaea GN=LEA3-3 PE=2 SV=1	35	29	19	10
E5FHY7	Late embryogenesis abundant protein group 3 protein OS=Arachis hypogaea GN=LEA3-5 PE=2 SV=1	47	33	23	12
E5FHY8	Late embryogenesis abundant protein group 3 protein OS=Arachis hypogaea GN=LEA3-7 PE=2 SV=1	12	4	13	8
E5FHY9	Late embryogenesis abundant protein group 4 protein OS=Arachis hypogaea GN=LEA4-1 PE=2 SV=1	23	9	14	6
E5FHZ0	Late embryogenesis abundant protein group 4 protein OS=Arachis hypogaea GN=LEA4-2 PE=2 SV=1	28	10	16	5
E5FHZ1	Late embryogenesis abundant protein group 5 protein OS=Arachis hypogaea GN=LEA5-1 PE=2 SV=1	14	0	5	0
E5FHZ2	Late embryogenesis abundant protein group 5 protein OS=Arachis hypogaea GN=LEA5-2 PE=2 SV=1	23	18	7	6
E5FHZ5	Late embryogenesis abundant protein group 7 protein OS=Arachis hypogaea GN=LEA7-1 PE=2 SV=1	14	12	12	8
E5FHZ6	Late embryogenesis abundant protein group 8 protein OS=Arachis hypogaea GN=LEA8-1 PE=2 SV=1	8	5	-	-
E5FHZ7	Late embryogenesis abundant protein group 9 protein OS=Arachis hypogaea GN=LEA9-1 PE=2 SV=1	5	4	3	2
E6Y6R4	Biotin carboxylase 1 OS=Arachis hypogaea GN=BC1 PE=2 SV=1	-	-	4	1

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
E6Y6R7	Biotin carboxylase 4 OS=Arachis hypogaea GN=BC4 PE=2 SV=1	-	-	2	0
E6Y6R8	Alpha-carboxyltransferase 1 OS=Arachis hypogaea GN=accA1 PE=2 SV=1	-	-	30	7
E6Y6R9	Alpha-carboxyltransferase 2 OS=Arachis hypogaea GN=accA2 PE=2 SV=1	-	-	26	5
E6Y9A7	Beta-ketoacyl-ACP synthetase I OS=Arachis hypogaea GN=kasI PE=2 SV=1	8	6	-	-
E6YC17	Symbiosis receptor kinase SymRK OS=Arachis hypogaea PE=1 SV=2	-	-	8	5
E7CQA1	Heat shock transcription factor A-2 OS=Arachis hypogaea subsp. hypogaea GN=HsfA2 PE=2 SV=1	-	-	7	5
E9JSA5	Putative acyl acyl-carrier-protein thioesterase type A OS=Arachis hypogaea GN=FATA PE=2 SV=1	-	-	6	6
E9JSA8	Acyl-[acyl-carrier-protein] desaturase OS=Arachis hypogaea GN=SAD1-1 PE=2 SV=1	-	-	1	0
F6KLJ6	Annexin OS=Arachis hypogaea var. vulgaris PE=2 SV=1	50	43	11	8
F8UX79	Glyceraldehyde-3-phosphate dehydrogenase OS=Arachis hypogaea PE=2 SV=1	7	2	-	-
G0UIN9	Cystatin OS=Arachis hypogaea GN=cst PE=4 SV=1	1	0	-	-
G0Y6T1	RNA-directed DNA polymerase OS=Arachis hypogaea GN=303L13_1 PE=4 SV=1	13	4	5	0
G0Y6T3	RNA-directed DNA polymerase OS=Arachis hypogaea GN=303L13_3 PE=4 SV=1	-	-	16	11
G0Y6T5	Putative uncharacterized protein OS=Arachis hypogaea GN=303L13_5 PE=4 SV=1	-	-	7	3
G0Y6T6	WD40-like protein OS=Arachis hypogaea GN=303L13_6 PE=4 SV=1	-	-	10	3
G0Y6T8	Putative uncharacterized protein OS=Arachis hypogaea GN=303L13_8 PE=4 SV=1	25	17	8	6
G0Y6T9	Putative uncharacterized protein OS=Arachis hypogaea GN=303L13_9 PE=4 SV=1	-	-	3	2

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
G0Y6U3	Putative uncharacterized protein OS=Arachis hypogaea GN=303L13_13 PE=4 SV=1	-	-	5	1
G0Y6U9	RNA-directed DNA polymerase OS=Arachis hypogaea GN=303L13_19 PE=4 SV=1	-	-	30	16
G0Y6V0	TIR-NBS-LRR-TIR type disease resistance protein OS=Arachis hypogaea GN=303L13_20 PE=4 SV=1	25	14	-	-
G0Y6V1	NBS-LRR type disease resistance protein OS=Arachis hypogaea GN=205D04_1 PE=4 SV=1	-	-	21	12
G0Y6V7	Transposon protein OS=Arachis hypogaea GN=205D04_7 PE=4 SV=1	-	-	25	12
G0Y6W1	TIR-NBS-LRR type disease resistance protein OS=Arachis hypogaea GN=205D04_11 PE=4 SV=1	52	26	51	28
G0Y6X0	Putative uncharacterized protein OS=Arachis hypogaea GN=205D04_20 PE=4 SV=1	-	-	19	15
G3F840	2-methyl-6-phytylbenzoquinone methyltransferase 1 OS=Arachis hypogaea GN=mpbqmt-1 PE=2 SV=1	-	-	4	2
G9HPX7	ADP-ribosylation factor OS=Arachis hypogaea GN=ARF PE=2 SV=1	4	3	7	4
G9HPX8	Peptidyl-prolyl cis-trans isomerase OS=Arachis hypogaea GN=CYC PE=2 SV=1	18	1	8	0
H6U596	Alcohol dehydrogenase (Fragment) OS=Arachis hypogaea PE=2 SV=1	9	9		
H6U598	Phosphatases pleiotropic regulator PRL1 (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	10	9
H6U599	Receptor serine/threonine kinase (Fragment) OS=Arachis hypogaea PE=2 SV=1	6	5	-	-
H6U5A1	Hypersensitive-induced response protein (Fragment) OS=Arachis hypogaea PE=2 SV=1	6	4	-	-
H9L792	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase OS=Arachis hypogaea GN=PGP1 PE=2 SV=1	6	3	5	3

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
I1E3D2	3-ketoacyl-CoA synthase OS=Arachis hypogaea GN=kcs PE=2 SV=1	-	-	9	7
J7LMM2	Eukaryotic translation initiation factor 5A OS=Arachis hypogaea GN=IF5A PE=2 SV=1	10	8	-	-
K4PM24	Absciscic acid 8'-hydroxylase CYP707A1 (Fragment) OS=Arachis hypogaea PE=2 SV=1	6	3	6	3
K7PQ65	Mother of FT and TFL1 OS=Arachis hypogaea GN=MFT PE=2 SV=1	-	-	2	1
K7PQ81	Flowering locus T OS=Arachis hypogaea GN=FT PE=2 SV=1	8	5	-	-
K7PQV2	Flowering locus T-like protein OS=Arachis hypogaea GN=FTL PE=2 SV=1	8	8	2	2
K7XD27	Stilbene synthase OS=Arachis hypogaea GN=STS PE=2 SV=1	12	0	-	-
L7PCI4	Small GTP binding protein OS=Arachis hypogaea GN=RabG3b PE=3 SV=1			2	2
M4IXA7	Beta-1,3-glucanase OS=Arachis hypogaea PE=2 SV=1	3	3		
M4TFZ7	Proteasome-associated protein ECM29 (Fragment) OS=Arachis hypogaea GN=ECM29 PE=2 SV=1	-	-	2	0
N1NF13	Uncharacterized protein OS=Arachis hypogaea GN=ARAX_AHF417E07-019 PE=4 SV=1	2	2	-	-
N1NJN2	Uncharacterized protein OS=Arachis hypogaea GN=ARAX_AHF417E07-008 PE=4 SV=1	-	-	15	10
N1NKG9	Seed maturation protein OS=Arachis hypogaea GN=ARAX_AHF417E07-003 PE=4 SV=1	19	5	9	3
O20356	Ribulose 1,5-bisphosphate carboxylase-oxygenase large subunit (Fragment) OS=Arachis hypogaea GN=rbcL PE=3 SV=1	51	0	31	0
P00330	Alcohol dehydrogenase 1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=ADH1 PE=1 SV=5	40	32	-	-

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
P00489	Glycogen phosphorylase, muscle form OS=Oryctolagus cuniculus GN=PYGM PE=1 SV=3	169	140	-	-
P00924	Enolase 1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=ENO1 PE=1 SV=3	42	29	-	-
P01066	Bowman-Birk type proteinase inhibitor A-II OS=Arachis hypogaea PE=1 SV=1	9	2	5	1
P01067	Bowman-Birk type proteinase inhibitor B-II OS=Arachis hypogaea PE=1 SV=1	2	2	-	-
P02769	Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4	106	91	-	-
P02872	Galactose-binding lectin OS=Arachis hypogaea PE=1 SV=3	24	0	17	0
P20178	Stilbene synthase 1 OS=Arachis hypogaea PE=2 SV=1	10	0	-	-
P20780	Arachin 21 kDa protein OS=Arachis hypogaea PE=1 SV=1	-	-	4	4
P51069	Stilbene synthase 3 OS=Arachis hypogaea PE=2 SV=1	15	0	-	-
P80926	33.0 kDa cold shock protein (Fragment) OS=Arachis hypogaea PE=1 SV=1	-	-	1	1
P86349	Elongation factor G, chloroplastic (Fragments) OS=Arachis hypogaea PE=1 SV=1	-	-	8	6
P92924	Putative uncharacterized protein (Fragment) OS=Arachis hypogaea PE=2 SV=1	1	0	2	0
Q06012	Endochitinase 1A (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	3	2
Q06H19	UDP-glucose pyrophosphorylase (Fragment) OS=Arachis hypogaea PE=2 SV=1	16	13	-	-
Q06H21	Ubiquitin/ribosomal protein S27a OS=Arachis hypogaea PE=2 SV=1	11	6	9	4
Q06H27	Tryptophan synthase alpha chain (Fragment) OS=Arachis hypogaea PE=2 SV=1	7	5	-	-
Q06H31	Translationally controlled tumor-like protein OS=Arachis hypogaea PE=2 SV=1	19	12	-	-

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
Q06H32	Thioredoxin fold OS=Arachis hypogaea PE=2 SV=1	12	5	2	2
Q06H34	Tonoplast intrinsic protein alpha TIP (Fragment) OS=Arachis hypogaea PE=2 SV=1	7	4	7	5
Q06H40	Polyubiquitin 10 (Fragment) OS=Arachis hypogaea PE=2 SV=1	5	0	6	0
Q0PCD5	NBS-LRR type disease resistance protein GnRGC3 (Fragment) OS=Arachis hypogaea PE=4 SV=1	-	-	6	3
Q0PCD6	NBS-LRR type disease resistance protein GnRGC1 (Fragment) OS=Arachis hypogaea PE=4 SV=1	-	-	4	3
Q0PKR5	Proteinase inhibitor (Fragment) OS=Arachis hypogaea GN=PI PE=2 SV=1	7	0	4	0
Q0Q0Q8	Type 4 metallothionein OS=Arachis hypogaea GN=MT4b PE=2 SV=1	9	7	3	3
Q0Q0Q9	Type 4 metallothionein OS=Arachis hypogaea GN=MT4a PE=2 SV=1	6	5	7	6
Q1HDS7	Superoxide dismutase [Cu-Zn] OS=Arachis hypogaea PE=2 SV=1	10	1	2	0
Q1PCR5	Proteasome subunit beta type (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	4	4
Q1PCR8	Glutaredoxin (Fragment) OS=Arachis hypogaea PE=2 SV=1	-	-	5	5
Q2HP01	Resistance protein Cre (Fragment) OS=Arachis hypogaea GN=CRE PE=4 SV=1	8	4	12	4
Q2HWT7	Phospholipase D OS=Arachis hypogaea GN=plda1 PE=2 SV=1	-	-	18	1
Q2HWT8	Phospholipase D OS=Arachis hypogaea GN=plda2 PE=2 SV=1	-	-	8	4
Q2KP14	Soluble diacylglycerol acyltransferase OS=Arachis hypogaea GN=DGAT3 PE=2 SV=1	-	-	7	3
Q2KPN4	Resistance protein PRG (Fragment) OS=Arachis hypogaea PE=4 SV=1	-	-	3	2
Q2KPP3	Resistance protein PRG (Fragment) OS=Arachis hypogaea PE=4 SV=1	-	-	2	0

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
Q2KPR8	Resistance protein PRG (Fragment) OS=Arachis hypogaea PE=4 SV=1	-	-	5	3
Q2KPR9	Resistance protein PRG (Fragment) OS=Arachis hypogaea PE=4 SV=1	-	-	4	1
Q2KPU0	Resistance protein PNTR (Fragment) OS=Arachis hypogaea PE=4 SV=1	-	-	1	1
Q2KPX9	Resistance protein PLTR (Fragment) OS=Arachis hypogaea PE=4 SV=1	-	-	2	1
Q2KQ02	Resistance protein PLTR (Fragment) OS=Arachis hypogaea PE=4 SV=1	-	-	5	5
Q2KQ09	Resistance protein PLTR (Fragment) OS=Arachis hypogaea PE=4 SV=1	-	-	3	1
Q2KQ10	Resistance protein PLTR (Fragment) OS=Arachis hypogaea PE=4 SV=1	8	4	2	0
Q2KQ18	Resistance protein PLTR (Fragment) OS=Arachis hypogaea PE=4 SV=1	10	3	1	0
Q2KQ68	Resistance protein PLTR (Fragment) OS=Arachis hypogaea PE=4 SV=1	6	3	2	2
Q2LDQ6	DREB2A-like protein OS=Arachis hypogaea PE=2 SV=1	-	-	6	5
Q2PK12	Actin depolymerizing factor-like protein OS=Arachis hypogaea PE=2 SV=1	8	6	-	-
Q2PXN2	17.5 kDa class I HSP (Fragment) OS=Arachis hypogaea PE=2 SV=1	9	7		
Q2PXN3	Disease resistance response protein (Fragment) OS=Arachis hypogaea PE=2 SV=1	9	8	5	5
Q2PXN4	Seed maturation protein LEA 4 (Fragment) OS=Arachis hypogaea PE=2 SV=1	20	3	8	4
Q2PXN9	Dessication protectant protein LEA 14 (Fragment) OS=Arachis hypogaea PE=2 SV=1	3	0	-	-
Q2PYP5	Putative transport protein Sec24C (Fragment) OS=Arachis hypogaea PE=2 SV=1	3	0	-	-
Q2VMU0	Serine protease inhibitor OS=Arachis hypogaea PE=2 SV=1	5	0	-	-
Q2VT95	Resveratrol synthase OS=Arachis hypogaea PE=2 SV=1	12	0	-	-

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
Q38711	Galactose-binding lectin (Fragment) OS=Arachis hypogaea GN=lec PE=2 SV=1	21	2	14	2
Q45W76	Ubiquitin-conjugating enzyme 2 OS=Arachis hypogaea PE=2 SV=1	10	5	7	3
Q45W78	Ubiquitin fusion protein OS=Arachis hypogaea PE=2 SV=1	16	10	16	10
Q45W80	Nucleoside diphosphate kinase OS=Arachis hypogaea PE=2 SV=1	11	7	-	-
Q45W82	Superoxide dismutase [Cu-Zn] OS=Arachis hypogaea PE=2 SV=1	11	2	4	1
Q45W85	Lea protein 1 OS=Arachis hypogaea PE=2 SV=1	16	1	22	3
Q4JME6	Lipoxygenase OS=Arachis hypogaea GN=Pnlox3 PE=2 SV=1	83	0	33	0
Q4JME7	Lipoxygenase OS=Arachis hypogaea GN=Pnlox2 PE=2 SV=1	97	6	44	4
Q4U4M1	LEA protein (Fragment) OS=Arachis hypogaea PE=2 SV=1	20	0	20	0
Q5I212	Putative root-knot resistance protein OS=Arachis hypogaea PE=2 SV=1	10	6	-	-
Q5QET5	Putative uncharacterized protein (Fragment) OS=Arachis hypogaea PE=2 SV=1	2	1	-	-
Q5XXQ4	PR10 protein OS=Arachis hypogaea GN=PR10 PE=2 SV=1	12	3	14	3
Q5XXY3	PR protein 4A (Fragment) OS=Arachis hypogaea PE=2 SV=1			2	2
Q647G3	Oleosin OS=Arachis hypogaea PE=2 SV=1	11	8	3	2
Q647G6	Cystatin OS=Arachis hypogaea GN=cst PE=2 SV=1	1	0	-	-
Q647H1	Conarachin OS=Arachis hypogaea PE=2 SV=1	130	106	76	60
Q6KBB0	Putative inorganic pyrophosphatase OS=Arachis hypogaea GN=pah767 PE=2 SV=1	-	-	3	0
Q6PWX0	Calmodulin OS=Arachis hypogaea GN=CaM-3 PE=4 SV=1	20	0	-	-

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
Q6R2U4	Calmodulin OS=Arachis hypogaea GN=CaM1 PE=4 SV=1	20	0	-	-
Q6R2U6	Calmodulin OS=Arachis hypogaea GN=CaM2 PE=2 SV=1	17	3	8	7
Q6R2U7	Calmodulin OS=Arachis hypogaea GN=CaM1 PE=2 SV=1	15	0	-	-
Q70KY0	9-cis-epoxycarotenoid dioxygenase OS=Arachis hypogaea GN=NCED1 PE=2 SV=1	-	-	2	1
Q7X973	Bowman-Birk trypsin inhibitor OS=Arachis hypogaea PE=2 SV=1	7	0	4	0
Q7Y0G7	Phospholipase D OS=Arachis hypogaea GN=PLD PE=2 SV=1	-	-	14	0
Q850G6	Putative lea 4 (Fragment) OS=Arachis hypogaea GN=Gsi-22 PE=2 SV=1	6	4	-	-
Q850G8	Putative lea 1 OS=Arachis hypogaea GN=Gsi-13 PE=2 SV=1	20	1	9	0
Q8GUS2	Trihydroxystilbene synthase OS=Arachis hypogaea PE=2 SV=1	11	0	-	-
Q9M5D3	Lipoxygenase OS=Arachis hypogaea PE=2 SV=1	82	7	35	3
Q9M5L4	Resveratrol synthase 3 OS=Arachis hypogaea GN=RS3 PE=3 SV=1	11	0	-	-
Q9SLV5	Resveratrol synthase OS=Arachis hypogaea PE=1 SV=1	8	0	-	-
Q9SW87	Acyl-[acyl-carrier-protein] desaturase OS=Arachis hypogaea PE=2 SV=1	-	-	4	1
T2B995	Malate dehydrogenase OS=Arachis hypogaea GN=MDH PE=2 SV=1	7	4	-	-
T2B9M0	Fructose-bisphosphate aldolase OS=Arachis hypogaea PE=2 SV=1	40	31	3	2
T2BAV2	Ribosome recycling factor OS=Arachis hypogaea GN=RRF PE=2 SV=1	-	-	9	8
U3MF97	40S ribosomal protein S23 OS=Arachis hypogaea PE=2 SV=1	3	1	-	-
U5Y4L6	Calmodulin OS=Arachis hypogaea GN=CAM PE=2 SV=1	20	0	-	-

UniProt Accession	Description	Raw peanut (samples 1, 2 and 3)		Roasted peanut flours (all batches)	
		Peptide count	Unique peptides	Peptide count	Unique peptides
V5T684	Putative R2R3 MYB protein 9 OS=Arachis hypogaea GN=MYB9 PE=4 SV=1	9	4	-	-
V5T688	Putative MYB-related protein 16 OS=Arachis hypogaea GN=MYB16 PE=4 SV=1	7	3	-	-
V5T697	Putative MYB-related protein 26 OS=Arachis hypogaea GN=MYB26 PE=4 SV=1	4	2	3	2
V5T6P8	Putative MYB-related protein 18 OS=Arachis hypogaea GN=MYB18 PE=4 SV=1	-	-	9	6
V5T7W6	Putative R2R3 MYB protein 8 OS=Arachis hypogaea GN=MYB8 PE=4 SV=1	-	-	5	5
V5T7W9	Putative MYB-related protein 14 OS=Arachis hypogaea GN=MYB14 PE=4 SV=1	-	-	2	1
V9PW61	Putative stilbene synthase (Fragment) OS=Arachis hypogaea PE=2 SV=1	7	0	-	-
V9PW62	Putative stilbene synthase (Fragment) OS=Arachis hypogaea PE=2 SV=1	7	0	-	-
W5QUZ4	Nonexpressor of pathogenesis-related protein 1 OS=Arachis hypogaea GN=NPR1 PE=2 SV=1	-	-	8	2
W5QUZ5	Nonexpressor of pathogenesis-related protein 1 OS=Arachis hypogaea GN=NPR1 PE=4 SV=1	-	-	10	6
W5QV08	Nonexpressor of pathogenesis-related protein 1 OS=Arachis hypogaea GN=NPR1 PE=2 SV=1	-	-	9	2
W5ZVL9	Resistance protein (Fragment) OS=Arachis hypogaea PE=4 SV=1	4	0	-	-
W5ZW67	Resistance protein (Fragment) OS=Arachis hypogaea PE=4 SV=1	4	0	-	-
X2C0H4	Cyclops1 (Fragment) OS=Arachis hypogaea GN=IPD3 PE=2 SV=1	10	5	14	9

REFERENCES

- (1) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. *Rapid Commun Mass Spectrom* **2004**, *18*, 2401-2414.
- (2) Chapman, M. D.; Pomes, A.; Breiteneder, H.; Ferreira, F. *J Allergy Clin Immunol* **2007**, *119*, 414-420.
- (3) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J Mol Biol* **1990**, *215*, 403-410.
- (4) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. *Bioinformatics* **2007**, *23*, 2947-2948.
- (5) Waterhouse, A. M.; Procter, J. B.; Martin, D. M.; Clamp, M.; Barton, G. J. *Bioinformatics* **2009**, *25*, 1189-1191.
- (6) Li, G. Z.; Vissers, J. P.; Silva, J. C.; Golick, D.; Gorenstein, M. V.; Geromanos, S. J. *Proteomics* **2009**, *9*, 1696-1719.
- (7) 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. *Mol Nutr Food Res* **2008**, *52 Suppl 2*, S272-285.