Plant Science 175 (2008) 321–329

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Proteomics analysis of mature seed of four peanut cultivars using two-dimensional gel electrophoresis reveals distinct differential expression of storage, anti-nutritional, and allergenic proteins

Kameswara Rao Kottapalli [a,b,1], Paxton Payton [b,1], Randeep Rakwal [c,d], Ganesh Kumar Agrawal [d], Junko Shibato [c], Mark Burow [a,e], Naveen Puppala [f,*]

[a] Texas Tech University, Department of Plant and Soil Science, Lubbock, TX 79409, United States
[b] United States Department of Agriculture Cropping Systems Research Laboratory, Lubbock, TX 79415, United States
[c] Health Technology Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba West, 16-1 Onogawa, Tsukuba 305-8569, Japan
[d] Research Laboratory for Biotechnology and Biochemistry, Kathmandu, Nepal
[e] Texas A&M University Agricultural Experiment Station, Lubbock, TX 79403, United States
[f] New Mexico State University Agricultural Science Center at Clovis, 2346 SR 288, Clovis, NM 88101, United States

ARTICLE INFO

Article history:
Received 24 January 2008
Received in revised form 12 April 2008
Accepted 13 May 2008
Available online 21 May 2008

Keywords:
Seed proteins
Allergens
Peanut market types
2D protein profiles
nESI-LC–MS/MS

ABSTRACT

Profiles of total seed proteins isolated from mature seeds of four peanut cultivars, New Mexico Valencia C (NM Valencia C), Tamspan 90, Georgia Green, and NC-7, were studied using two-dimensional gel electrophoresis coupled with nano-electrospray ionization liquid chromatography tandem mass spectrometry (nESI-LC–MS/MS). Two-dimensional gels stained with silver nitrate revealed a total of 457, 516, 556, and 530 protein spots in NM Valencia C, Tamspan 90, Georgia Green, and NC-7, respectively. Twenty abundant protein spots showing differences in relative abundance among these cultivars were analyzed by nESI-LC–MS/MS, resulting in identification of 14 non-redundant proteins. The majority of these proteins belonged to the globulin fraction consisting of arachin (glycinin and Arah3/4) and conarachin seed storage proteins as well as other allergen proteins. The expression of some of these identified protein spots was cultivar-specific. For example, allergen Arah3/Arah4 and conarachin protein spots were only detected in Tamspan 90 and NC-7, whereas the Gly1 protein spot was detected only in NM Valencia C and NC-7. Moreover, a galactose-binding lectin protein spot with anti-nutritive properties was only present in Tamspan 90. Other proteins showing differences in relative abundance among the four cultivars included 13-lipoxygenase, fructose-biphosphate aldolase, and glyceraldehyde 3-phosphate dehydrogenase. Together, these results suggest that identified proteins might serve as potential markers for cultivar differentiation and may be associated with underlying sensory and nutritional traits of peanut cultivars.

Published by Elsevier Ireland Ltd.

1. Introduction

Peanut (*Arachis hypogaea* L.), an important crop legume with approximately 24% protein in the seed [1] is the third-most important source of plant protein, contributing 11% of world's protein supply. Peanut is primarily grown as a seed crop in the United States, but represents a rich source of edible fats and a primary source of oil in many parts of the world. In the United States, cultivated peanuts are mainly classified into four major market types: runner, Virginia, Spanish, and Valencia. The runner market type predominates production in the Southeast (Georgia, Florida, Alabama, and South Carolina) and Southwest (Texas,

* Corresponding author. Tel.: +1 505 985 2292; fax: +1 505 985 2419.
E-mail address: npuppala@nmsu.edu (N. Puppala).
[1] These authors equally contributed to the development of this manuscript.

Abbreviations: AMBIC, ammonium bicarbonate; CBB, coomassie brilliant blue; DTT, dithiothreitol; IPG, immobilized protein gradient; IEF, isoelectric focusing; iTRAQ, multiplexed isobaric tagging technology; LB-TT, lysis buffer containing thiourea and tris; MS, mass spectrometry; 2-ME, 2-mercaptoethanol; MudPIT, multi-dimensional protein identification technology; NCBI, National Center for Biotechnology Information; nESI-LC–MS/MS, nano-electrospray ionization liquid chromatography tandem mass spectrometry; 1-DGE, one-dimensional gel electrophoresis; PMF, peptide mass fingerprinting; RT, room temperature; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; 2-DGE, two-dimensional gel electrophoresis.

0168-9452/$ – see front matter. Published by Elsevier Ireland Ltd.
doi:10.1016/j.plantsci.2008.05.005

Oklahoma, and New Mexico). The Virginia market types are grown primarily in Virginia and North Carolina. Although runner peanuts comprise the majority of peanut acreage in the Southwest, significant quantities of Spanish and Valencia are also cultivated in the region. Runner and Virginia market types belong to subspecies *hypogaea* var. *hypogaea* whereas the Valencia and Spanish market type belongs to ssp. *fastigiata* var. *fastigiata* and ssp. *fastigiata* var. *vulgaris*, respectively. At present, the runner market type occupies 84% of the total peanut acreage in the United States followed by Virginia (13%), Spanish (2%), and Valencia (1%) market types [2]. Among runner market types, Georgia Green is the most widely grown cultivar because of its resistance to tomato spotted wilt virus. This single cultivar occupies 66% of the runner market and 55% of the US market. The cultivar NC-7, a Virginia type, occupies 6.4% of the Virginia market and less than 10% of the total US market. The Spanish-type cultivar Tamspan 90, occupies 64% of the Spanish market and about 2% of the US market. Valencia peanuts, NM Valencia A and C, constitute a niche crop for eastern New Mexico and are known for their superior taste and three or more seeds per pod. Recent efforts to identify novel molecular markers to distinguish the market types have been reported. We classified all the four market types in the US peanut mini-core collection using 12 mapped SSR markers [3], and Liang et al. [4] employed a protein-based marker approach. However, there remain very few systematic studies differentiating peanut cultivars from different market types based on molecular markers.

Numerous studies conducted during the last several decades on peanut seed proteins emphasized storage protein composition [5–7], nutritional value [8], and immunological properties of the protein fractions [9–11]. However, metabolic pathways and mechanisms of differential accumulation of various protein components that result in differences in seed morphology and quality (taste, allergenicity) remain largely unknown. A better understanding of the function of proteins expressed in mature seeds might help not only in cultivar differentiation, but also in improving the nutritional value of the peanut seed. Transcriptional profiling of developing *Arabidopsis* seeds has provided insight into the primary transcriptional networks that coordinate the genome-wide response to seed developmental programs and lead to the distribution of carbon among oil, protein, and carbohydrate reserves [12,13]. Cellular signaling and metabolic events are also driven by protein–protein interactions, post-translational protein modifications, and enzymatic activities that cannot be predicted accurately or described by transcriptional profiling approaches alone. Since proteins are directly associated with function, proteomics approaches are being applied increasingly to address biochemical and physiological questions in many model species such as *Arabidopsis*, *Medicago*, and *Brassica*. Recently, a combined transcriptomic and proteomics analyses of microspore-derived embryo development identified candidate embryo markers and proteins involved in protein biosynthesis, glycolysis, and ascorbate metabolism in rapeseed [14]. Indeed in the past few years, two-dimensional gel electrophoresis (2-DGE)-based proteomics approaches have been applied systematically to identify and profile proteins expressed during seed development or in the mature seed of model plant species like soybean [15], rapeseed [16], *Medicago* [17], *Arabidopsis* [18], wheat [19–21], and barley [22]. Although a few preliminary studies using 2-DGE were reported [4,23], to date there is no systematic proteomics-level study on peanuts as a whole and specifically on peanut seed using high-resolution 2-DGE for cultivar differentiation.

Here, we have optimized a 2-DGE-based proteomics approach for peanut seed as a first step toward our goal to establish comprehensive seed proteomes of different peanut cultivars. In this study, a 2-DGE-based proteomics approach was applied to four peanut cultivars: New Mexico Valencia C, Tamspan 90, Georgia Green, and NC-7. Out of 20 protein spots differing in relative abundance (either distinctly present and/or absent or quantitatively variable) among the four cultivars, 14 non-redundant proteins were identified by nano-electrospray ionization liquid chromatography tandem mass spectrometry (nESI-LC–MS/MS). The majority of these proteins belonged to the globulin fraction of seed proteins. We believe that these proteins could be used to distinguish these popular peanut cultivars representing the four US peanut market types. It is also likely that some of these proteins may be associated with the basic differences in sensory attributes and nutritional traits including allergens in these cultivars.

2. Materials and methods

2.1. Seed material

Four peanut cultivars, NM Valencia C, Tamspan 90, Georgia Green and NC-7, representing the four major market types, were used in this study (Fig. 1). All seeds were obtained from plants grown at the Agricultural Experiment Station at Clovis, New Mexico. Pods were harvested at full maturity and air-dried seeds were stored at 4 °C before use.

2.2. Peanut seed protein extraction using phenol

Six randomly selected, fully matured dry seeds (for each of the four cultivars) were ground to a fine powder in liquid nitrogen. Approximately 100 mg of homogenized material was used for protein extraction with phenol. The standardized protocol is detailed in Supplementary Fig. 1. Briefly, the homogenized sample

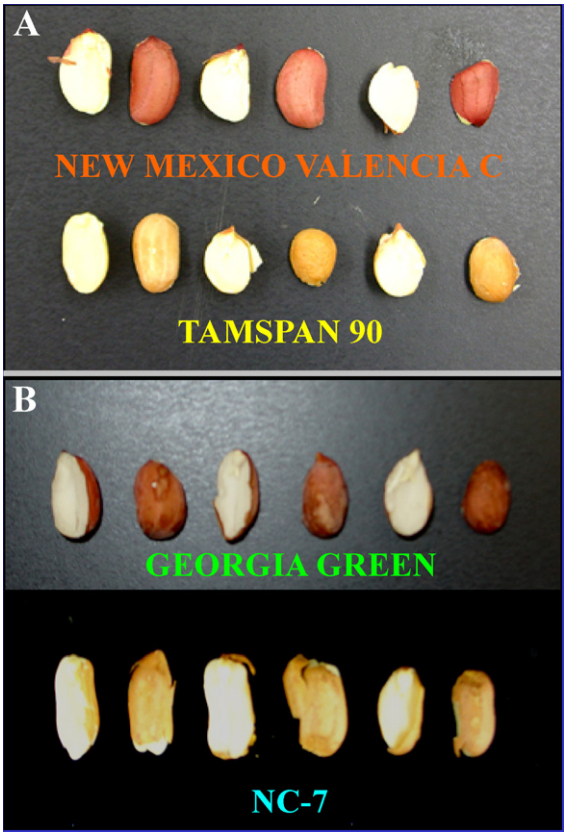

Fig. 1. Morphological differences in the mature seeds of four peanut cultivars, NM Valencia C, Tamspan 90, Georgia Green, and NC-7.

was gently mixed with extraction media [0.9 M sucrose, 0.1 M Tris–HCl (pH 8.8), 10 mM EDTA (pH 8.0) and 0.4% (v/v) 2-mercaptoethanol (2-ME) in Milli Q (MQ) water] and Tris-buffered phenol (pH 8.8) at room temperature (RT). Proteins were precipitated by incubating the phenol phase with 0.1 M ammonium acetate–methanol at −20 °C overnight, followed by precipitation and washing of the proteins serially in three organic solvents to give a highly purified protein pellet. The protein pellet was then used for both 1-DGE (presented in Supplementary Fig. 2) and 2-DGE (described below).

To compare our phenol protein extraction method to a previously published protocol (hexane/TCA–acetone) [4], protein was extracted from mature seeds of Tamspan 90 by both methods and separated by 2-DGE. The comparative 2D protein profile of Tamspan 90 clearly indicated that the phenol extraction method resulted in enhanced resolution and separation of extracted soluble proteins on 2D gels (data not shown). We concluded that phenol extraction was as good as hexane/TCA–acetone method, but the clear separation of the spots in low-, mid- and high-molecular weight regions was far superior in the phenol method.

2.3. Two-dimensional gel electrophoresis

The protein pellet was solubilized in LB-TT [7 M Urea, 2 M thiourea, 4% (w/v) CHAPS, 18 mM Tris–HCl (pH 8.0–8.3), 14 mM Trizma base, 0.2% (v/v) Triton X-100, EDTA-free proteinase inhibitor cocktail tablet (Roche Diagnostics GmbH, Mannheim, Germany) and 50 mM dithiothreitol (DTT) in autoclaved MQ water]. After centrifugation at 12,000 × *g* for 15 min (10 °C), the protein in the supernatant was quantified using a Coomassie Plus™ protein assay kit (PIERCE, Rockford, IL), and stored in aliquots at −80 °C. The 2-DGE was carried out using pre-cast IPG strip (pH 4–7; 24 cm) gels on an IPGphor unit followed by the second dimension using pre-cast ExcelGel™ XL SDS 12–14 (gradient) on a Multiphor II horizontal electrophoresis unit. All these gels and instruments were from GE Healthcare Bio-Sciences AB (Uppsala, Sweden). The volume carrying 80 μg total soluble protein was mixed with LB-TT containing 0.5% (v/v) pH 4–7 IPG buffer to a final volume of 450 μL. A trace of bromophenol blue was added and centrifuged at 12,000 × *g* for 15 min followed by pipetting into a 24-cm strip holder tray placed into the IPGphor unit. IPG strips were carefully placed onto the protein samples, avoiding air bubbles between the sample and the gel. The IPG strips were allowed to passively rehydrate with the protein samples for 1.5 h, followed by overlaying the IPG strips with cover fluid (mineral oil), and this was directly linked to a five-step active rehydration and focusing protocol (24 cm strip) as described previously [24]. The entire procedure was performed at 20 °C, and a total of 76908 Vh was used for isoelectric focusing (IEF).

The strip gels were incubated in equilibration buffer (50 mM Tris–HCl (pH 8.8), 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS) containing 2% (w/v) DTT for 10 min (twice) with gentle agitation, followed by incubation in the same equilibration buffer supplemented with 2.5% (w/v) iodoacetamide for the same time periods as above at RT. SDS-PAGE was performed as per the recommendations in the Multiphor II manual. For each sample, three IPG strips and corresponding SDS-PAGE runs were used under the same conditions. To visualize the protein spots, polyacrylamide gels were stained with silver nitrate (Plus One Silver Staining Kit Protein; GE Healthcare). Protein patterns in the gels were recorded as digitalized images using a Canon CanoScan 8000F digital scanner (resolution 300 dpi), and saved as TIFF files. Gels were quantitated in profile mode using ImageMaster 2D Platinum software version 5.0 (GE Healthcare). Differential spots were defined as those showing a 1.5-fold or greater change (increase/decrease) in spot volumes across all three gel replications. These results (spot differences between the four peanut genotypes) were also confirmed manually. Moreover, silver nitrate-stained spots were selected for comparative profiling only if they were confirmed in all three independent gel replications.

2.4. In-gel trypsin digestion

Protein spots were excised from a 2D gel using a gel picker (One Touch Spot Picker, P2D1.5, The Gel Company, San Francisco, CA) and transferred into sterile 1.5 ml micro-centrifuge tubes. In-gel digestion with sequencing-grade modified trypsin was carried out according to a previous procedure [25] with minor modifications. Briefly, protein spots were washed twice with 100 mM ammonium bicarbonate (pH 8.5, hereafter called AMBIC) and then dehydrated with acetonitrile. Gel pieces were reduced with 10 mM DTT at 56 °C for 45 min and alkylated with 50 mM iodoacetamide for 45 min at RT in the dark in AMBIC solution. Gel pieces were then washed with 20 mM AMBIC, dehydrated with acetonitrile, and air dried. The dried gel pieces were subjected to in-gel digestion with 20 μL of 20 mM AMBIC containing 15 ng/μL sequence grade modified trypsin (17,000 U/mg; Promega, Madison, WI) at 37 °C for 18 h. The peptides were extracted twice with 20 μL of 20 mM AMBIC and thrice with 20 μL of 0.5% trifluoroacetic acid (TFA) in 50% acetonitrile. The peptides extracted in the five steps were combined and concentrated to 20 μL using a centrifugal concentrator (CC-105; TOMY, Tokyo, Japan).

2.5. Identification of 2D gel protein spots by nano-electrospray ionization liquid chromatography mass spectrometry (nESI-LC–MS/MS)

The peptides (20 μL) were used for mass spectral analysis on a LCQ Deca linear ion trap mass spectrometer (nESI-LC–MS/MS; Thermo Electron) through a nano-electrospray ionization source. Briefly, the on-line capillary LC included a monolithic reverse-phase trap column (0.2 mm × 5 cm, MonoCap for fast-flow, GL Science, Tokyo, Japan) and a fast-equilibrating C18 capillary column (monolith-type column; i.d, 0.1 mm; length, 50 mm; GL Science). Sample was loaded onto peptide traps for concentration and desalting prior to final separation by the C18 column using a linear acetonitrile gradient ranging from 5% to 65% solvent B [H_2O/acetonitrile/formic acid, 10/90/0.1 (v/v)] in solvent A [H_2O/acetonitrile/formic acid, 98/2/0.1 (v/v)] for a duration of 40 min. The mass/charge (m/z) ratios of eluted peptides and fragmented ions from a fused-silica Fortis Tip emitter (150 μM o.d., 20 μM i.d.; AMR Inc., Tokyo, Japan) were analyzed in the data-dependent positive acquisition mode on LC–MS/MS. Dynamic exclusion parameters used were repeat count (2), repeat duration (0.5 min), exclusion list size (25), and exclusion duration (3.0 min). Following each full scan (400–2000 m/z), a data-dependent triggered MS/MS scan for the most intense parent ion was acquired. The heated fused-silica Fortis Tip emitter was held at 1.8 kV for ion sprays and a flow rate 300 nL/min.

2.6. Database search, data analysis, and protein identification

Xcalibur Version 1.1 software (Thermo Fisher Scientific K.K.) was used for generating the peaklist. Acquired data were searched against the National Center for Biotechnology Information (NCBInr) plant protein database (nrdb90, update 2006.09.09) for plant sequences using the MASCOT software package (Version 2.1, Matrix Science, Inc., London, U.K.). The NCBInr database contained 3,947,950 sequences and 1,358,419,857 residues, and the taxonomic class searched was Viridiplantae (Green plants; 268,880

Table 1
Peanut seed protein identification by nESI-LC–MS/MS

Spot no.	Protein name	Accession	MW (Da)/p*I* (theoretical)	MW (kDa)/p*I* (observed)	Spot relative abundance (cvs. 1/2/3/4)[a]	Score	Sequence coverage (%)	Matched amino acid sequence	MASCOT ion score
1	13-Lipoxygenase (*Arachis hypogaea*)	gi|68161356	97820/5.37	96.6/5.66	0.18/0.36/0.37/0.27	111	15	YREEDLK, GDGKGER, NDPNSEKPGDVYVPR, LPTEVISTISPLPVIK, TDGEQVLK, EMIAGVNPCMIR, RINETHAK, EGVESTIWLLAK, QNLINSDGIIER, GMAVEDSSSPYGIR, LLPEQGTAEYEEMVK, and GVPNSISI	29,4,53,62,9,25,23, 12,39,10,55,12
2	13-Lipoxygenase (*A. hypogaea*)	gi|68161356	97820/5.37	92.1/5.55	0.19/0.38/0.39/0.21	118	16	GTVVLMR, YREEDLK, NDPNSEKPGDVYVPR, SSDFLANSIK, LPTEVISTISPLPVIK, TDGEQVLK, RINETHAK, TILFLK, DTMNINALAR, QNLINSDGIIER, TFLPSK, GMAVEDSSSPYGIR, LLPEQGTAEYEEMVK, and GVPNSISI	10,9,25,42,7,63,16,22,23,19, 28,33,59,45,10,28,51,22,17,8
4	Conarachin (*A. hypogaea*)	gi|46560476	48122/5.43	61.1/5.12	0.17/0.18/0.19/0.30	42	5	IVQIEAKPNTLVLPK, and AMVIVVVNK	42,6
5	Conarachin (*A. hypogaea*)	gi|52001225	76343/5.23	50.8/5.06	0.00/0.24/0.00/0.28	95	16	IESQGGITETWNSNHPELR, CAGVTLLK, ESEQEQEQQGDSHHK, ESEQEQEQQGDSHHKIYHFR, RFYLAGNPEEEHPETQQQQPQTR, ILQNPPEQTKDQIVR, VEGGFRDVISPR, DVISPR, and QGYDDDDRRPR	60,19,30,16,50,38,24,28,18,9
8	Fructose-bisphosphate aldolase, cytoplasmic isozyme	gi|3913008	38599/6.21	34.4/6.76	0.20/0.38/0.21/0.34	105	14	YHDELIANAAYIGTPGK, GILAADESTGTIGKR, VAPEVVAEHTVR, and ALQQSTLK	98,19,39,9
9	Glyceraldehyde 3-phosphate dehydrogenase (*Torenia hybrida*)	gi|34850871	21488/7.93	31.9/6.77	0.22/0.40/0.26/0.44	145	23	LTGMAFR, VPTVDVSVVDLTVR, GILGYTEDDVVSTDFVGDSR, and AGIALSK	7,82,92,25
10	Allergen Arah3/Arah4 (*A. hypogaea*)	gi|21314465	62042/5.52	34.2/5.71	0.00/0.64/0.00/0.16	349	20	QQPEENACQFQR, RPFYSNAPQEIFIQQGR, GYFGLIFPGCPSTYEEPAQQGR, RHQSQRPPR, RFNLAGNHEQEFLR, FNLAGNHEQEFLR, FNLAGNHEQEFLR, SLPYSPYSPQTQPK, QILQNLRGENESDEQGAIVTVR, and GENESDEQGAIVTVR	42,28,53,13,49,69,28,41,34,72
11	Gly1 (*A. hypogaea*)	gi|9864777	60754/5.48	33.9/5.48	0.27/0.00/0.00/0.40	218	21	QQPEENACQFQR, RPFYSNAPQEIFIQQGR, GYFGLIFPGCPSTYEEPAQQGR, RFNLAGNHEQEFLR, FNLAGNHEQEFLR, SLPYSPYSPQSQPR, QEEREFSPR, QIVQNLR, GENESEEEGAIVTVK, and ILSPDR	17,36,45,33,55,2418,18,5812
12	Glycinin (*A. hypogaea*)	gi|47933675	58368/5.41	29.5/5.09	0.45/0.68/0.58/0.64	440	29	QGGEENECQFQR, IESEGGYIETWNPNNQEFQCAGVALSR, RFQVGQDDPSQQQQDSHQK, FQVGQDDPSQQQQDSHQK, RFYLAGNQEQEFLR, FYLAGNQEQEFLR, YQQQQGSRPHYR, VRGDEQENEGSNIFSGFAQEFLQHAFQVDR, GDEQENEGSNIFSGFAQEFLQHAFQVDR, GENEREEQGAIVTVK, ILSPDEEDESSR, and REEFDEDR	45,46,40,80,19,26,5, 43,31,34,73,25
13	Glycinin (*A. hypogaea*)	gi|5712199	61430/5.48	27.9/4.62	0.23/0.14/0.15/0.24	99	8	QQPEENACQFQR, RPFYSNAPQEIFIQQGR, and RFNLAGNHEQEFLR	21,67,46
14	Os01g0290100 (*Oryza sativa* (japonica cultivar))	gi|115436056	52430/6.12	26.9/6.53	0.32/0.49/0.30/0.47	26	6	AGATVVEVPLPFPVASADAIIAEFRAALDVAK	26
17	Galactose-binding lectin precursor (agglutinin) (PNA)	gi|1708792	29364/5.39	24.0/5.15	0.25/0.46/0.00/0.00	32	13	VLYAMPVR, TVPWNSVSGAVVK, VTVIYDSSTK, and AKLPER	32,16,18,9
18	Galactose-binding lectin precursor (agglutinin) (PNA)	gi|1708792	29364/5.39	23.3/5.01	0.00/0.42/0.00/0.01	57	16	VLYAMPVR, TVPWNSVSGAVVK, VTVIYDSSTK, and SWSFTSTLITTTR	40,21,34,51

Table 1 (*Continued*)

Spot no.	Protein name	Accession	MW (Da)/p*I* (theoretical)	MW (kDa)/p*I* (observed)	Spot relative abundance (cvs. 1/2/3/4)[a]	Score	Sequence coverage (%)	Matched amino acid sequence	MASCOT ion score
20	Glycinin (*A. hypogaea*)	gi|47933675	58368/5.41	18.0/6.37	0.26/0.36/0.37/0.30	819	37	VRGDEQENEGSNIFSGFAQEFLQHAFQVDR, GDEQENEGSNIFSGFAQEFLQHAFQVDR, SSNPDIYNPQAGSLR, SVNELDLPILGWLGLSAQHGTIYR, NAMFVPHYTLNAHTIVVALNGR, AHVQVVDSNGNR, AHVQVVDSNGNRVYDEELQEGHVLVVPQNFAVAAK, VYDEELQEGHVLVVPQNFAVAAK, AQSENYEYLAFK, TDSRPSIANLAGENSIIDNLPEEVVANSYR, NNNPFK, and FFVPPFDHQSMR	41,67,54,61,82,42,41, 37,55,65,13,6

[a] % intensity was calculated by ImageMaster software; cvs. 1–4 correspond to New Mexico Valencia C, Tamspan 90, Georgia Green, and NC-7.

sequences). The typical parameters used in the MASCOT MS/MS ion search were maximum of one trypsin missed cleavage, fixed modification cysteine carbamidomethylation, variable modification methionine oxidation, peptide mass tolerance ±2 Da, threshold ($p < 0.05$), minimum ion counts (0), and fragment mass tolerance ±0.8 Da. A minimum of two peptides were considered as positive hits for protein identification. The cutoff score/expectation value for accepting individual MS/MS spectra was 25, which indicates identity or extensive homology of probability lower than 0.05. The threshold employed was $-10 \times \log(P)$, where P is the probability that the observed match is a random event; protein scores are derived from ion scores as a non-probabilistic basis for ranking protein hits. When peptides matched to multiple members of a protein family, the unique peptide of identified proteins were considered. The unique peptide of each protein was manually selected to reduce the multi-matching peptide mistake to regenerate the protein identification list. For unambiguous identification of isoforms/individual members of a protein family, the experimental molecular mass and p*I* of the protein spot was taken into account. Single non-redundant proteins were assigned for multiple proteins ID with the same accession number. The peptide sequences were listed along with the corresponding proteins.

3. Results

3.1. 1-DGE analysis of total seed proteins

Given the distinct morphological differences between the four cultivars, we hypothesized that 1-DGE would differentiate the cultivars based on broad protein level differences. 1-DGE on 12.5% SDS-PAGE revealed a high molecular weight protein band which suggests that the total protein extracted by our standardized protocol was intact. However, no distinct differences between cultivars were found on SDS-PAGE (Supplementary Fig. 2).

3.2. 2-DGE analysis of peanut seed reveals distinct differences in protein profiles of NM Valencia C, Tamspan 90, Georgia Green, and NC-7

Varietal differences of protein profiles in mature peanut seeds were investigated utilizing 2-DGE analysis. Representative silver nitrate-stained 2D gels are shown in Fig. 2A and B. The 2D gels revealed 457, 516, 556, and 530 protein spots in NM Valencia C, Tamspan 90, Georgia Green, and NC-7, respectively. This is the first report of high-resolution 2-DGE in peanuts using an optimized phenol extraction protocol in protein profiling studies. ImageMaster analysis identified a total of 20 abundant protein spots that changed markedly among the 4 cultivars. The nESI-LC–MS/MS analysis of these protein spots resulted in identification of 14 non-redundant proteins (Table 1). The enlarged gel regions having nine protein spots representing six individual proteins are shown in Fig. 3. Protein spots 1, 2, 4, and 19 quantitatively differed in expression while protein spots 5, 11, 10, 17, and 18 were present in some but absent in other cultivars. Protein spot 19 with unknown function was quantitatively variable among the cultivars and is a candidate for future characterization studies. Major classes of differential proteins identified are discussed below.

3.3. Storage proteins

3.3.1. Conarachin

Protein spots 4 and 5 matched a conarachin protein, a highly abundant peanut storage protein. Protein spot 4 was highest in NC-7 (1.8-fold higher than NM Valencia C and Georgia Green and 2-fold higher than Tamspan 90). In contrast, protein spot 5 was

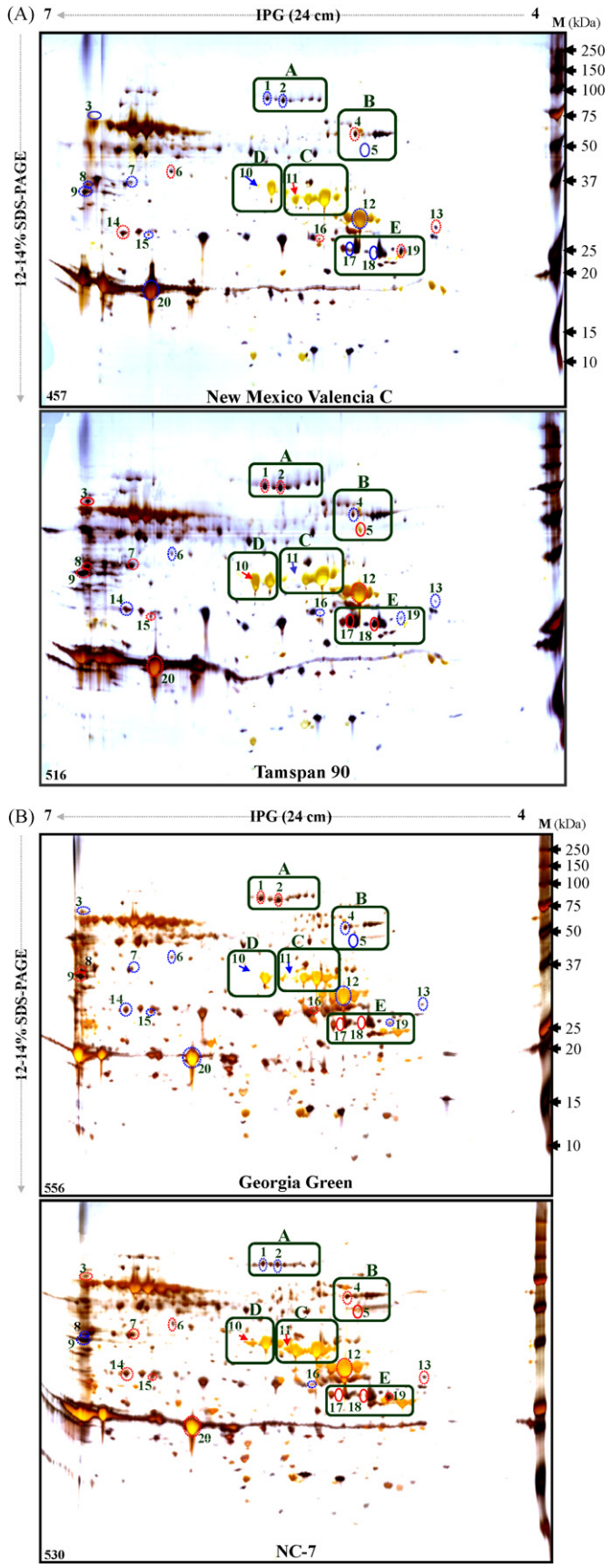

Fig. 2. Seed protein profiles of four peanut cultivars. A, NM Valencia C and Tamspan 90, and B, Georgia Green and NC-7. Seed proteins (80 μg) were separated on 24 cm IPG (pH 4–7) strips in the first dimension followed by SDS-PAGE on gradient (12–14%) large-format gels in the second dimension. Proteins were stained with silver nitrate. 5 μL of the molecular mass standards (Bio-Rad) were loaded adjacent to the acidic end of the IPG strip. Image analysis detected 457, 516, 556 and 530 spots on the 2D gels, which are indicated on the bottom left-hand corners. Numbered arrows or circles (spots 1–20) show proteins that were present (or increased) or absent (or decreased) in all four gels. These spots runner are color-coded, with red arrows or circles showing presence and the blue arrows or circles showing absence of the spots. The broken circles represent increase or decrease in the spots. The positions of the proteins are marked in the same direction and position for clarity, in each of the gels. The rectangular boxes indicate the areas showing major differences among these cultivars. Enlarged forms of these rectangular boxes is shown in Fig. 3.

absent in New Mexico Valencia C and Georgia Green but present in Tamspan-90 and NC-7 cultivars.

3.3.2. Globulin proteins—Gly1 and glycinin

Four protein fractions of varying mass (spots 11, 12, 13, and 20) and corresponding to glycinin protein could differentiate the four peanut cultivars. Interestingly protein spot 11 (Fig. 3) was not detected in Tamspan 90 and Georgia Green cultivars, but was expressed 1.5-fold higher in NC-7 compared to NM Valencia C. Protein spots 12, 13 and 20 were variably expressed among the 4 cultivars.

3.4. Allergen proteins

3.4.1. Arah3/Arah4

Allergen Arah3/Arah4 (protein spot 10) was completely absent in New Mexico Valencia C and Georgia Green cultivars. Tamspan 90 cultivar had the highest levels (fourfold higher than NC-7 cultivar) of Arah3/h4.

3.5. Anti-nutritive and off-flavor related proteins

3.5.1. Galactose-binding lectin

Galactose-binding lectin (spots 17 and 18) was found highly expressed in Tamspan 90 but absent in NM Valencia C, Georgia Green and NC-7 cultivars.

3.5.2. Lipoxygenase (LOX)

Two protein spots (spots 1 and 2) that were differentially expressed in the four peanut cultivars were identified as 13-LOX of 98 kDa molecular mass. The 13-LOX protein was highly expressed (2.5-fold higher) in Tamspan 90 compared to NM Valencia C (sweetest cultivar). Georgia Green and NC-7, cultivars with intermediate taste [26], showed moderate levels of the protein (2- and 1.5-fold higher than NM Valencia C).

3.6. Enzymes of glycolysis

Two differential protein spots (spots 8 and 9) identified were involved in glycolysis and gluconeogenesis/the Calvin cycle. Protein spot 8 was identified as fructose-bisphosphate aldolase cytoplasmic isozyme and spot 9 is identical to glyceraldehyde-3-phosphate dehydrogenase, a tetrameric NAD-binding enzyme. These two proteins were highly expressed in Tamspan 90 and NC-7 but were reduced (1.8-fold) in NM Valencia C and Georgia Green cultivars.

4. Discussion

Liang et al. [4] identified differences on isoforms of basic arachin (iso-*Ara* h3) among runner and Spanish types. Their results suggest that diversity in this group of iso-allergens may be the product of post-translation modification. Moreover, their work provided insight toward identifying polymorphic seed proteins as

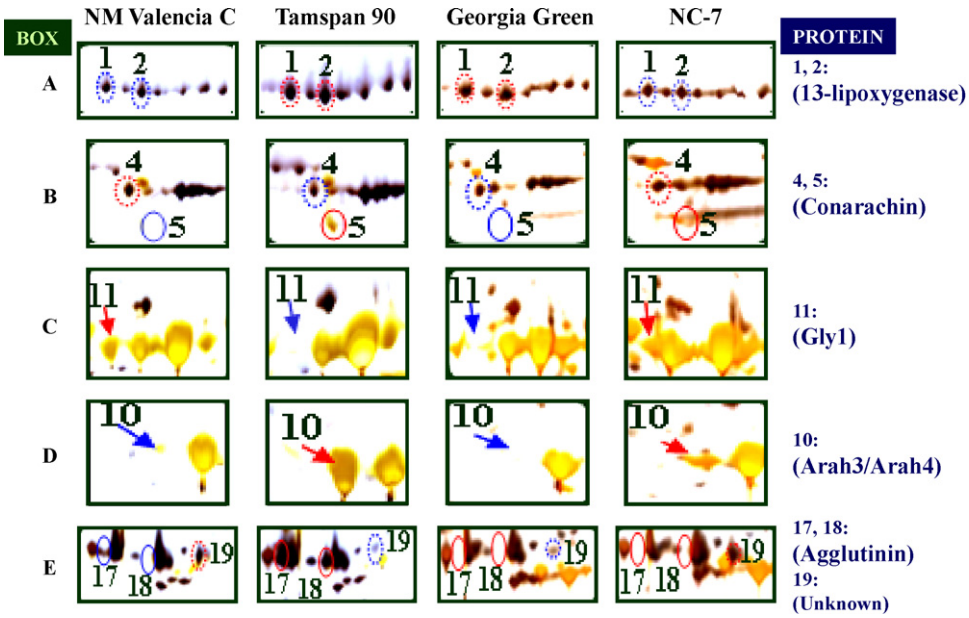

Fig. 3. Dynamic variations in seed proteins of four peanut cultivars. Five different rectangular gel regions (regions A–E) derived from Fig. 2A and B were selected to display the dynamic variations of protein spot abundance in these cultivars. The p*I* and molecular masses of these protein spots varies between 5–6 and 25–100 kDa, respectively. Assigned proteins corresponding to protein spots are given on the right-hand side.

markers for cultivar identification. To expand that hypothesis, we employed an optimized phenol protein extraction method for 2-DGE in order to identify additional marker proteins in all four major US market types. A second objective was to identify proteins responsible for the diverse sensory attributes seen among theses cultivars. A significant number of distinct morphological differences exist among ssp. *fastigiata* and *hypogaea*. Subspecies *fastigiata*, represented in this study by NM Valencia C and Tamspan 90, are early maturing with an erect growth habit, large, yellow-green pigmented leaves, alternate branching, and have mainstem flowers. Valencia peanuts typically have three or more seeds per pod, and Spanish market types have pods with two small seeds. Georgia Green and NC-7, belonging to ssp. *hypogaea*, are generally late maturing with a recumbent growth habit, small, dark-green leaves, no mainstem flowers, and show a sequential branching pattern. This subspecies is characterized by having two seeds per pod, with Virginia types having large-sized seeds, and runner types producing medium seeds. The two subspecies also differ in absence or presence of seed dormancy, another potentially interesting characteristic that may be elucidated through seed proteomics. Roasted peanut sensory attributes of the four peanut cultivars were evaluated by Pattee et al., who distinguished them based on roasted peanut, sweet, and bitter traits [26]. According to their study, NM Valencia C had the highest score for sweet attribute of any line tested. The NC-7 cultivar had the lowest score for sweet and highest score for bitter attributes. Tamspan 90 was inferior in the sweet attribute compared to NM Valencia C but superior to Georgia Green.

Although total proteins extracted were intact on 12.5% SDS-PAGE similar to other studies using 15% gels [4], no differences were noticed suggesting that 1-DGE could not be used for differentiating the selected cultivars. These results are similar to previous 1-DGE studies that have shown limited or no variation between cultivated peanuts [4,27,28]. Our optimized phenol extraction and resolution by 2-DGE, however, enabled the identification of 20 distinct protein spots differentially expressed among the four cultivars and their identification by LC–MS/MS. Not surprisingly, these proteins belonged to the two major functional categories, storage (including allergens) and anti-nutritive lectin proteins, that make up the majority of seed protein content [29].

Differential expression of peanut allergens and storage proteins was observed among the four cultivars. Glycinin, an 11S-type seed storage globulin protein encoded by peanut *Gly1* [30] showed significant variation among the cultivars (Fig. 3). Glycinin is nutritionally superior to the 7S con-glycinins [7], and possesses superior intrinsic functional properties for processed foods [31]. Additionally, immunological studies demonstrated that glycinin-type proteins are less allergenic compared to vicilin- and conglutin-type seed storage proteins [8,9]. Allergen Arah3/Arah4 is an isoform reported to have decreased allergenicity [11] and was found abundantly in Tamspan 90, accumulated to moderate levels in NC-7, but was present a very low levels in NM Valencia C and not detected in Georgia Green. This cultivar needs to be further investigated and its Arah3/h4 content quantified, before it can qualify as parent for breeding hypoallergenic peanuts.

Phytohemagglutinins, a major legume seed lectins have a unique property of binding carbohydrate substances and play a role in the packaging and transport of storage materials [32]. However, the presence of lectins may have an adverse effect on the nutritional quality of raw legume seeds [33,34] and has been implicated in the proliferation of colon cancer cells by interaction with glycosylated transmembrane glycoprotein isoforms [35]. Additionally, lectin content has a direct effect on protein digestibility and low lectin levels are a desired seed trait [36]. Galactose-binding lectin accumulated to significant levels in Tamspan 90 but was undetectable in mature seeds of the other cultivars indicating their superior seed proteins quality.

All four cultivars also expressed various levels of LOX proteins in the mature seed. LOX proteins catalyze the oxygenation at the C-13 site of polyunsaturated fatty acids such as linoleic (C18:2) or linolenic acids (C18:3). LOX proteins are important commercially [37,38] because their activity destroys essential polyunsaturated fatty acids and gives rise to compounds that directly contribute to peanut sensory attributes, namely taste and odor [31,39], and shelf-life [40]. In soybean, the development of off-flavor is most likely due to production of short chain volatile carbonyl

compounds that bind to soya proteins. In pea and lentil, LOX activity is responsible for the development of rancid off-flavor [40–42]. Given the multifunctional role of lipoxygenase, it is likely that LOX identified in this study might be involved in seed quality such as taste/flavor in peanuts. We expect that lipoxygenase could be associated with rancid flavors as lower levels of LOX were noticed in NM Val C, a cultivar known for superior taste and shelf-life.

Decreased expression of fructose-bisphosphate aldolase cytoplasmic isozyme and glyceraldehyde-3-phosphate dehydrogenase, the two proteins involved in glycolysis may suggest a general reduction in energy levels. Although the exact function of these two proteins in seeds is largely unknown, but nevertheless, they are polymorphic across the market types and serve as markers for cultivar differentiation.

Apparent differences in the expression of identified proteins in these four cultivars represents a valuable tool for cultivar identification in breeding studies and expands our knowledge of the peanut seed proteome. Furthermore, protein profiling in these four peanut cultivars and their identification implies that no single cultivar has all desirable traits for breeding a cultivar to increase seed quality. For example, New Mexico Valencia C showed low levels of the anti-nutritive proteins LOX and galactose-binding lectin. Conversely, Arah3/h4, an allergen with decreased allergenic properties was highly abundant in Tamspan 90. Since there are marked differences between the four cultivars with respect to flavor attributes, some of these proteins, LOX in particular, could be associated with rancidity and off-flavor. It would be interesting to know whether the identified protein differences can be correlated to the actual market type variations by generating proteome maps of several cultivars from each distinct market type. Further studies are needed to comprehensively investigate the peanut seed proteome by both gel- and non-gel-based proteomics approaches in a quantitative manner such as multiplexed isobaric tagging technology (iTRAQ) and multi-dimensional protein identification technology (MudPIT) analysis.

Acknowledgements

We thank Ms. Misato Hirano (HTRC) and undergraduate students Mr. Kiyotaka Horie and Ms. Takako Furusawa (School of Agriculture, Ibaraki University, Ami, Japan) for their technical support. The expert assistance of Ms. Yuko Ogawa (HTRC) in MS analysis is also appreciated. We thank Drs. Hitoshi Iwahashi, Yoshinori Masuo, and Yasukazu Yoshida, of HTRC, for their generous support and encouragement. This research was supported by grants from the National Peanut Board, Texas Peanut Producers Board, New Mexico Peanut Research Board, New Mexico State University Agricultural Experiment Station, Ogallala Aquifer Initiative, USDA-ARS CRIS 6208-21000-012-00D, and USAID-Peanut CRSP through University of Georgia.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.plantsci.2008.05.005.

References

[1] J.P. Cherry, Potential sources of peanut seed proteins and oil in the genus *Arachis*, J. Agric. Food Chem. 25 (1976) 186–193.
[2] AOSCA, Association of Official Seed Certifying Agencies, 2005, http://aosca.org/index.htm.
[3] K.R. Kottapalli, G. Burow, J. Burke, M.D. Burow, N. Puppala, Molecular characterization of the U.S. peanut minicore collection using microsatellite marker, Crop Sci. 47 (2007) 1718–1727.
[4] X.Q. Liang, M. Luo, C.C. Holbrook, B.Z. Guo, Storage protein profiles in Spanish and runner market type peanuts and potential markers, BMC Plant Biol. 6 (2006) 24.
[5] D.B. Jones, M.H. Horn, The properties of arachin and con-arachin and the proportionate occurrence in the peanut, J. Agric. Res. 40 (1930) 672–682.
[6] P. Johnson, E.M. Shooter, E.H. Rideal, The globulins of the groundnut. II. Electrophoretic examination of the arachin system, Biochim. Biophys. Acta 5 (1950) 376–395.
[7] J.P. Cherry, J.M. Dechary, R.L. Ory, Gel electrophoretic analysis of peanut proteins and enzymes. 1. Characterization of DEAE-cellulose separated fractions, J. Agric. Food Chem. 21 (1973) 652–655.
[8] A. Millerd, Biochemistry of legume seed proteins, Ann. Rev. Plant Physiol. 26 (1975) 53–72.
[9] O.M. Viquez, K.N. Konan, H.W. Dodo, Structure and organization of the genomic clone of a major peanut allergen gene, *Ara h1*, Mol. Immunol. 40 (2003) 565–571.
[10] S.J. Koppelman, M. Wensing, M. Ertmann, A.C. Knulst, E.F. Knol, Relevance of *Ara h1* Ara h2 and Ara h3 in peanut-allergic patients, as determined by immunoglobulin E western blotting, basophil-histamine release and intracutaneous testing: Ara h2 is the most important peanut allergen, Clin. Exp. Allergy 34 (2004) 583–590.
[11] I.H. Kang, M. Gallo, Cloning and characterization of a novel peanut allergen Ara h3 isoform displaying potentially decreased allergenicity, Plant Sci. 172 (2007) 345–353.
[12] T. Girke, J. Todd, S. Ruuska, J. White, C. Benning, J. Ohlrogge, Microarray analysis of developing *Arabidopsis* seeds, Plant Physiol. 124 (2000) 1570–1581.
[13] S.A. Ruuska, T. Girke, C. Benning, J.B. Ohlrogge, Contrapuntal networks of gene expression during *Arabidopsis* seed filling, Plant Cell 14 (2002) 1191–1206.
[14] R. Joosen, J. Cordewener, E.D. Supena, O. Vorst, M. Lammers, C. Maliepaard, T. Zeilmaker, B. Miki, T. America, J. Custers, K. Boutilier, Combined transcriptome and proteome analysis identifies pathways and markers associated with the establishment of rapeseed microspore-derived embryo development, Plant Physiol. 144 (2007) 155–172.
[15] M. Hajduch, A. Ganapathy, J.W. Stein, J.J. Thelen, A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database, Plant Physiol. 137 (2005) 1397–1419.
[16] M. Hajduch, J.E. Casteel, K.E. Hurrelmeyer, Z. Song, G.K. Agrawal, J.J. Thelen, Proteomic analysis of seed filling in Brassica napus. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis, Plant Physiol. 141 (2006) 32–46.
[17] K. Gallardo, C. Le Signor, J. Vandekerckhove, R.D. Thompson, J. Burstin, Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation, Plant Physiol. 133 (2003) 664–682.
[18] K. Gallardo, C. Job, S.P.C. Groot, M. Puype, H. Demol, J. Vandekerckhove, D. Job, Proteomics of *Arabidopsis* seed germination. A comparative study of wild-type and gibberellin-deficient seeds, Plant Physiol. 129 (2002) 823–837.
[19] D.J. Skylas, L. Copeland, W.G. Rathmell, C.W. Wrigley, The wheat-grain proteome as a basis for more efficient cultivar identification, Proteomics 1 (2001) 1542–1546.
[20] N. Islam, S.H. Woo, H. Tsujimoto, H. Kawasaki, H. Hirano, Proteome approaches to characterize seed storage proteins related to ditelocentric chromosomes in common wheat (*Triticum aestivum* L.), Proteomics 2 (2002) 1146–1155.
[21] T. Majoul, E. Bancel, E. Triboi, J. Ben Hamida, G. Branlard, Proteomic analysis of the effect of heat stress on hexaploid wheat grain: characterization of heat-responsive proteins from total endosperm, Proteomics 3 (2003) 175–183.
[22] C. Finnie, K. Maeda, O. Østergaard, K.S. Bak-Jensen, J. Larsen, B. Svensson, Aspects of the barley seed proteome during development and germination, Biochem. Soc. Trans. 32 (2004) 517–519.
[23] A. Boldt, D. Fortunato, A. Conti, A. Petersen, B. Ballmer-Weber, U. Lepp, G. Reese, W.M. Becker, Analysis of the composition of an immunoglobulin E reactive high molecular weight protein complex of peanut extract containing Ara h1 and Ara h3/4, Proteomics 5 (2005) 675–686.
[24] G.K. Agrawal, N.S. Jwa, Y. Iwahashi, M. Yonekura, H. Iwahashi, R. Rakwal, Rejuvenating rice proteomics: facts, challenges, and vision, Proteomics 6 (2006) 5549–5576.
[25] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, Anal. Chem. 68 (1996) 850–858.
[26] H.E. Pattee, T.G. Isleib, D.W. Gorbet, F.G. Giesbrecht, Z. Cui, Parental selection in breeding for roasted peanut flavor quality, Peanut Sci. 28 (2001) 51–58.
[27] A.K. Singh, S. Sivaramakrishnan, M.H. Mengesha, C.D. Ramaiah, Phylogenetic relations in section Arachis based on seed protein profile, Theor. Appl. Genet. 82 (1991) 593–597.
[28] A.K. Singh, S. Gurtu, R. Jambunathan, Phylogenetic relationships in the genus *Arachis* based on seed protein profiles, Euphytica 74 (1994) 219–225.
[29] A. Vitale, R. Bollini, Legume storage proteins, in: J. Kigel, G. Galili (Eds.), Seed Development and Germination, Marcel Dekker, New York, USA, 1995 pp. 73–102.
[30] A.K. Jain, Cloning and structural analysis of a cDNA clone encoding glycinin (*Gly-1*) seed storage protein of peanut, Electron. J. Biotechnol. 7 (2004) 189–199.
[31] J.E. Kinsella, Functional properties of soybean proteins, J. Am. Oil Chem. Soc. 56 (1979) 242–258.
[32] M.E. Etzler, Distribution and function of plant lectins, in: I.E. Liener, N. Sharon, I.J. Goldstein (Eds.), The Lectins, Academic Press, New York, 1986, pp. 371–435.

[33] I.E. Liener, Phytohemagglutinins (phytolectins), Ann. Rev. Plant Physiol. 27 (1976) 291–319.
[34] I.E. Liener, Factors affecting the nutritional quality of soya products, J. Am. Oil Chem. Soc. 58 (1981) 406–415.
[35] R. Singh, S. Subramanian, J.M. Rhodes, B.J. Campbell, Peanut lectin stimulates proliferation of colon cancer cells by interaction with glycosylated CD44v6 isoforms and consequential activation of c-Met and MAPK: functional implications for disease-associated glycosylation changes, Glycobiology 16 (2006) 594–601.
[36] R. Bollini, E. Carnovale, B. Campion, Removal of antinutritional factors from bean (*Phaseolus vulgaris* L.) seeds, Biotechnol. Agron. Soc. Environ. 3 (1999) 217–219.
[37] T. Galliard, H.W.S. Chan, in: E.E. Conn, P.K. Stumpf (Eds.), The Biochemistry of Plants, vol. 4, Academic Press, London/New York, 1980, pp. 131–161.
[38] J. Nicolas, M. Autran, R. Drapron, Purification and some properties of wheat germ lipoxygenase, J. Sci. Food Agric. 33 (1982) 365–372.
[39] A. Pinsky, S. Grossman, M. Trop, Lipoxygenase content and antioxidant activity of some fruits and vegetables, J. Food Sci. 36 (1971) 571–572.
[40] R. Casey, C. Domoney, N.C. Nielsen, Isolation of a cDNA clone for pea (*Pisum sativum*) seed lipoxygenase, Biochem. J. 232 (1985) 79–85.
[41] B. Bengtsson, I. Bosund, Lipid hydrolysis in unblanched frozen peas (*Pisum sativum*), J. Food Sci. 31 (1966) 474–481.
[42] N.A.M. Eskin, H.M. Henderson, Ann. Technol. Agric. 26 (1977) 139–149.