A Systematic Proteomic Study of Seed Filling in Soybean. Establishment of High-Resolution Two-Dimensional Reference Maps, Expression Profiles, and an Interactive Proteome Database[1][w]

Martin Hajduch, Ashwin Ganapathy, Joel W. Stein, and Jay J. Thelen*

Department of Biochemistry, Life Sciences Center (M.H., J.J.T.), and Computer Science Department (A.G., J.W.S.), University of Missouri, Columbia, Missouri 65211

A high-throughput proteomic approach was employed to determine the expression profile and identity of hundreds of proteins during seed filling in soybean (*Glycine max*) cv Maverick. Soybean seed proteins were analyzed at 2, 3, 4, 5, and 6 weeks after flowering using two-dimensional gel electrophoresis and matrix-assisted laser desorption ionization time-of-flight mass spectrometry. This led to the establishment of high-resolution proteome reference maps, expression profiles of 679 spots, and corresponding matrix-assisted laser desorption ionization time-of-flight mass spectrometry spectra for each spot. Database searching with these spectra resulted in the identification of 422 proteins representing 216 nonredundant proteins. These proteins were classified into 14 major functional categories. Proteins involved in metabolism, protein destination and storage, metabolite transport, and disease/defense were the most abundant. For each functional category, a composite expression profile is presented to gain insight into legume seed physiology and the general regulation of proteins associated with each functional class. Using this approach, an overall decrease in metabolism-related proteins versus an increase in proteins associated with destination and storage was observed during seed filling. The accumulation of unknown proteins, sucrose transport and cleavage enzymes, cysteine and methionine biosynthesis enzymes, 14-3-3-like proteins, lipoxygenases, storage proteins, and allergenic proteins during seed filling is also discussed. A user-intuitive database (http://oilseedproteomics.missouri.edu) was developed to access these data for soybean and other oilseeds currently being investigated.

Soybeans are responsible for approximately $12 billion in annual crop value and more than $5 billion in annual export value to the U.S. economy (Gunstone, 2001). Soybeans (and other oilseeds) are a significant source of fatty acids and proteins for human and animal nutrition as well as for nonedible uses, including industrial feedstocks and combustible fuel (for review, see Thelen and Ohlrogge, 2002). The major source of these commodities is the seed, and a better understanding of this structure will be important for future biotechnology efforts.

During a 4- to 5-week period of soybean seed development, most of the storage reserves for germination are synthesized; this phase of development is generally termed seed filling. At maturation, approximately 41% of soybean seed dry weight is storage protein (Hill and Breidenbach, 1974), and the two prevalent seed storage proteins are glycinin and β-conglycinin (Roberts and Briggs, 1965; Hill and Breidenbach, 1974; Thanh and Shibasaki, 1976). Storage proteins accumulate primarily in the protein storage vacuoles of terminally differentiated cells of the embryo and endosperm and as protein bodies directly assembled within endoplasmic reticulum (Herman and Larkins, 1999). The synthesis of storage proteins and the formation of specialized vacuoles occur after cell division is complete, when all further growth occurs only through cell expansion and accumulation of storage reserves (for review, see Herman and Larkins, 1999). Approximately 18% to 21% of soybean seed dry weight is oil in the form of triacylglycerol. From 24 to 40 d after flowering (DAF), oil percentage increases rapidly and by the end of this period accounts for approximately 30% of the total oil of the mature seed. The remaining 70% is synthesizing during 40 to 64 DAF, also a period of seed desiccation (Rubel et al., 1972).

Seed filling has been investigated recently in Arabidopsis (*Arabidopsis thaliana*) on a global, transcriptome level. Microarray analyses of developing seeds revealed complex networks of gene expression (Girke et al., 2000; Ruuska et al., 2002). For example, a number of genes encoding core fatty acid synthesis enzymes displayed a bell-shaped pattern of expression between 5 and 13 DAF. By contrast, the expression of storage proteins, oleosins, and other known abscisic acid-regulated genes increased later and remained high. With recent development in protein identification

[1] This work was supported by the National Science Foundation Plant Genome Research Program (Young Investigator Award no. DBI–0332418).

* Corresponding author; e-mail thelenj@missouri.edu; fax 573–884–9676.

[w] The online version of this article contains Web-only data.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.104.056614.

techniques, large-scale profiling of anonymous proteins is becoming routine, and these analyses can also reveal information about posttranslation modifications in addition to protein accumulation. However, despite the importance of seed filling, systematic proteomic analysis of this phase of seed development is only beginning to be investigated in legumes. Proteomics of seed development was investigated recently in *Medicago truncatula* (Gallardo et al., 2003). This study confirmed the identity and expression of proteins previously proposed to accumulate during seed development and also provided information about proteins presumably involved in cell division during embryogenesis.

Other proteomics investigations of legume seeds include two surveys of the principal proteins expressed in mature soybean seeds (Herman et al., 2003; Mooney and Thelen, 2004). These investigations identified approximately 100 and 40 proteins, respectively, from mature soybean seeds and provided the basis for future comparative analysis of soybean seed development. These investigations also demonstrated that high-throughput proteomics can be performed on soybean using either peptide mass fingerprinting (PMF) or tandem mass spectrometry for protein identification. As the complete genome sequence for this crop presently is not available, databases of nonredundant, contiguous expressed sequence tags were relied upon in both investigations for bioinformatic mining. In this investigation, we present a systematic study of more than 600 proteins expressed during five key stages of seed development in soybean. A Web-based database was developed to archive all expression and PMF data for public accession in a user-intuitive format.

RESULTS

Staging and Characterization of Developing Soybean Seed

The primary objective of this study was to characterize global protein expression during the seed-filling phase of soybean seed development. For the best coverage of this period, whole seeds were analyzed at precisely 2, 3, 4, 5, and 6 weeks after flowering (WAF). The experimental period included the late morphogenic phase (2 WAF), which is completed when seeds are about 2 mm long, the period of cell division (3, 4 WAF), and the cell enlargement period (5, 6 WAF) but not the early embryogenesis or the seed maturation phases (Mienke et al., 1981). Figure 1 shows the characteristics of the developing seeds used in this study. These data are in agreement with measurements published previously (Rubel et al., 1972; Yazdi-Samadi et al., 1977). However, fresh and dry weight values are slightly higher, which can be attributed to differences between varieties and growth conditions.

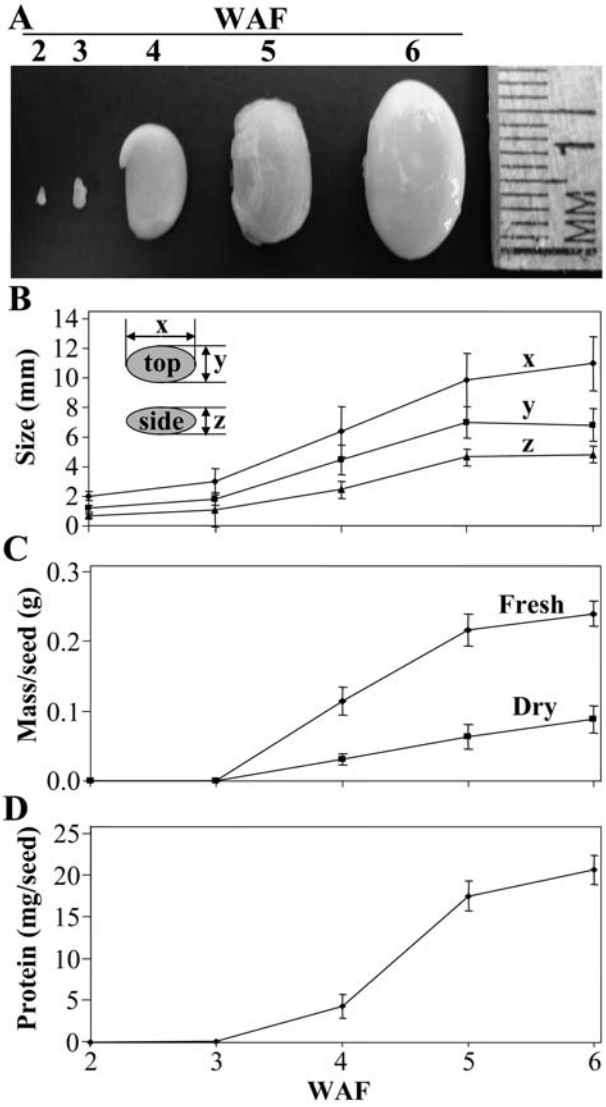

Figure 1. Development of soybean seeds during the experimental period. A, Whole seeds at the five stages of seed filling. Experimental sampling began exactly 2 WAF and continued at precisely 7-d intervals until 6 WAF. B, Individual seed size characteristics, including length (x), thickness (y), and width (z), were determined using an ultramicrometer. Each value is the average of 10 seeds; SD is denoted by error bars. C, Seed fresh and dry mass during the experimental period expressed as mass per seed. Values are the average of 10 determinations; SD is shown. D, Total protein content per seed during the investigated period of seed filling. Values are the average of 10 determinations; SD is shown.

Narrow-Range Isoelectric Focusing Is Necessary for High-Resolution Proteome Maps

Whole proteins from staged developing soybean seed were resolved and detected using high-resolution two-dimensional electrophoresis (2-DE) followed by colloidal Coomassie Blue staining. Initial analyses were performed with immobilized pH gradient (IPG) strips that ranged from pH 3 to 10 (Fig. 2A). It was observed that the region from pH 4 to 7 was a highly dense area on the proteome map; therefore, additional

analyses with pH 4 to 7 IPG strips were performed to improve spot resolution (Fig. 2B). The 2-DE maps showed a highly dynamic proteome during soybean seed development. The late morphogenesis phase of seed development (2 WAF) and the early cell division stage (3 WAF) showed similar 2-DE spot patterns. Late cell division (4 WAF) and cell enlargement (5 and 6 WAF) periods were characterized by an increasing abundance of seed storage proteins, which accounted for approximately 35%, 53%, and 60% of total seed protein, respectively.

Altogether 679 Spot Groups Were Quantified Using 2-DE

After 2-DE, gels were imaged and analyzed using ImageMaster Platinum software (Amersham Biosciences, Piscataway, NJ). For protein expression analyses, the volume of each spot was expressed as relative volume, a ratio of individual spot volume to the sum of spot volumes for all analyzed spots. Analysis of relative volume, instead of raw volumes, corrected for subtle experimental variations due to protein loading and staining. Moreover, the relative volumes were

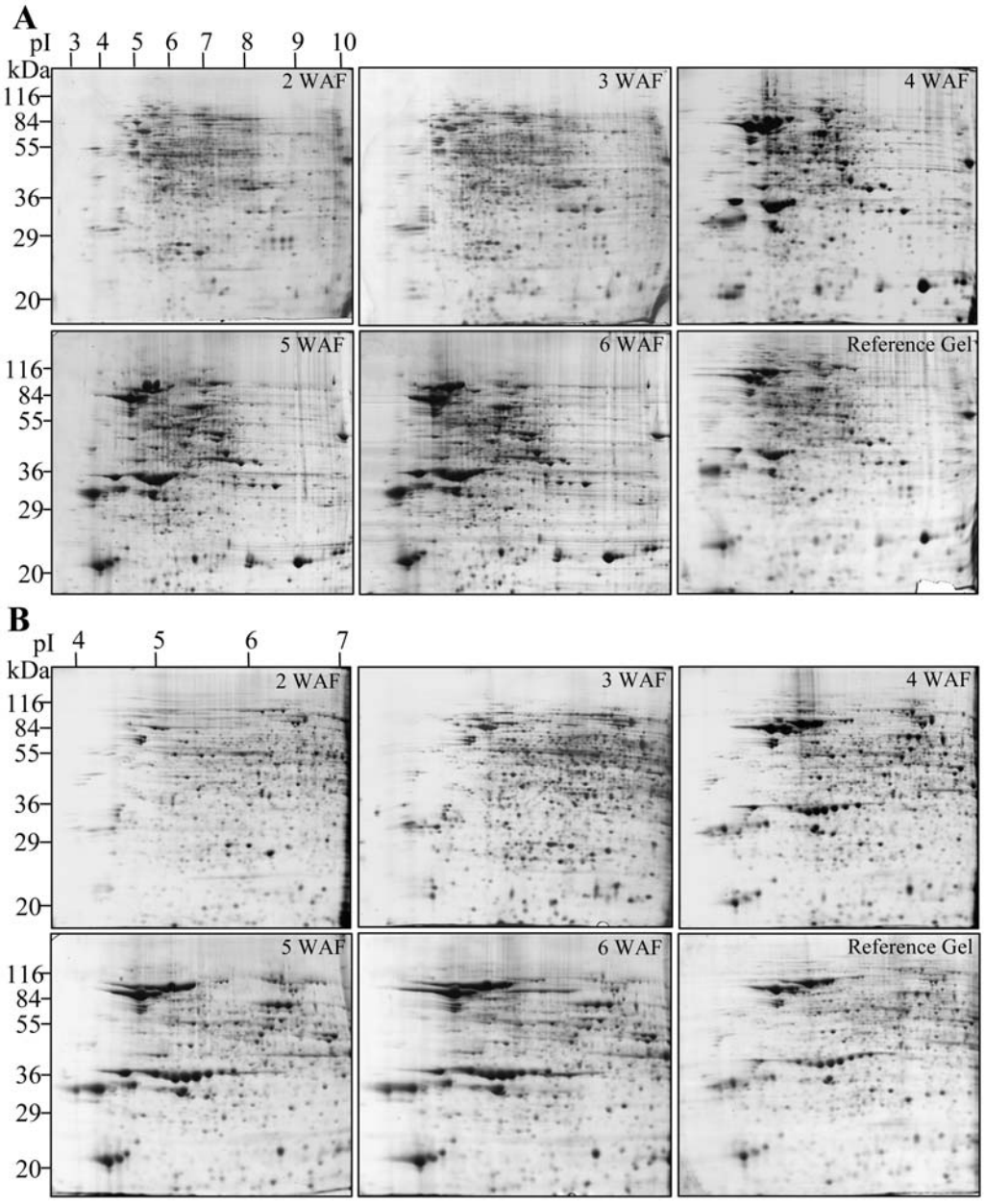

Figure 2. Analysis of proteins (1 mg) isolated from immature soybean seeds of 2-, 3-, 4-, 5-, and 6-WAF periods by 2-DE in combination with colloidal Coomassie Brilliant Blue staining. A, Protein analysis using wide-range IPG strips with pH range from 3 to 10. B, Two-dimensional electrophoresis of highly dense region of pH 4 to 7 using narrow-range IPG strips. Isoelectric points (pI) and molecular mass (in kD) are noted. Reference gel is a composite of all 5 seed stages obtained by pooling 0.2 mg of protein from each stage.

adjusted with correction constants to enable direct comparison between the pI 4 to 7 and pI 3 to 10 datasets. The relative volumes for a particular spot, obtained from at least three biological analyses, were averaged. Within these protein groups, the protein expression data and SD values for each developmental stage were plotted on a line graph (Fig. 3).

To ensure the quality of expression data, threshold criteria were established for image analysis. To be included in expression analyses, each spot group needed to be represented in at least three out of four biological replicates. In total, 488 proteins satisfied these criteria from 2-DE gels of pH range 4 to 7. An additional 191 proteins from pH 3 to 10 gels (pH 7–10 region only) also satisfied these criteria (Fig. 4). The software used for this analysis assigned a unique number to each detected spot within each gel in the series (spot 3514 in Fig. 3). After analyzed gels were matched to a reference gel, protein groups were established and assigned a spot group number (1100 in Fig. 3) corresponding to the number of that particular spot on the reference gel. Using this systematic approach, the expression profiles for 679 spot groups were obtained in biological triplicate for each of the five developmental stages.

Peptide Mass Fingerprinting Using the Soybean UniGene Database Yielded 422 Protein Assignments

Each of the 679 proteins with confirmed expression profiles was excised from reference gels for identification by matrix-assisted laser desorption ionization

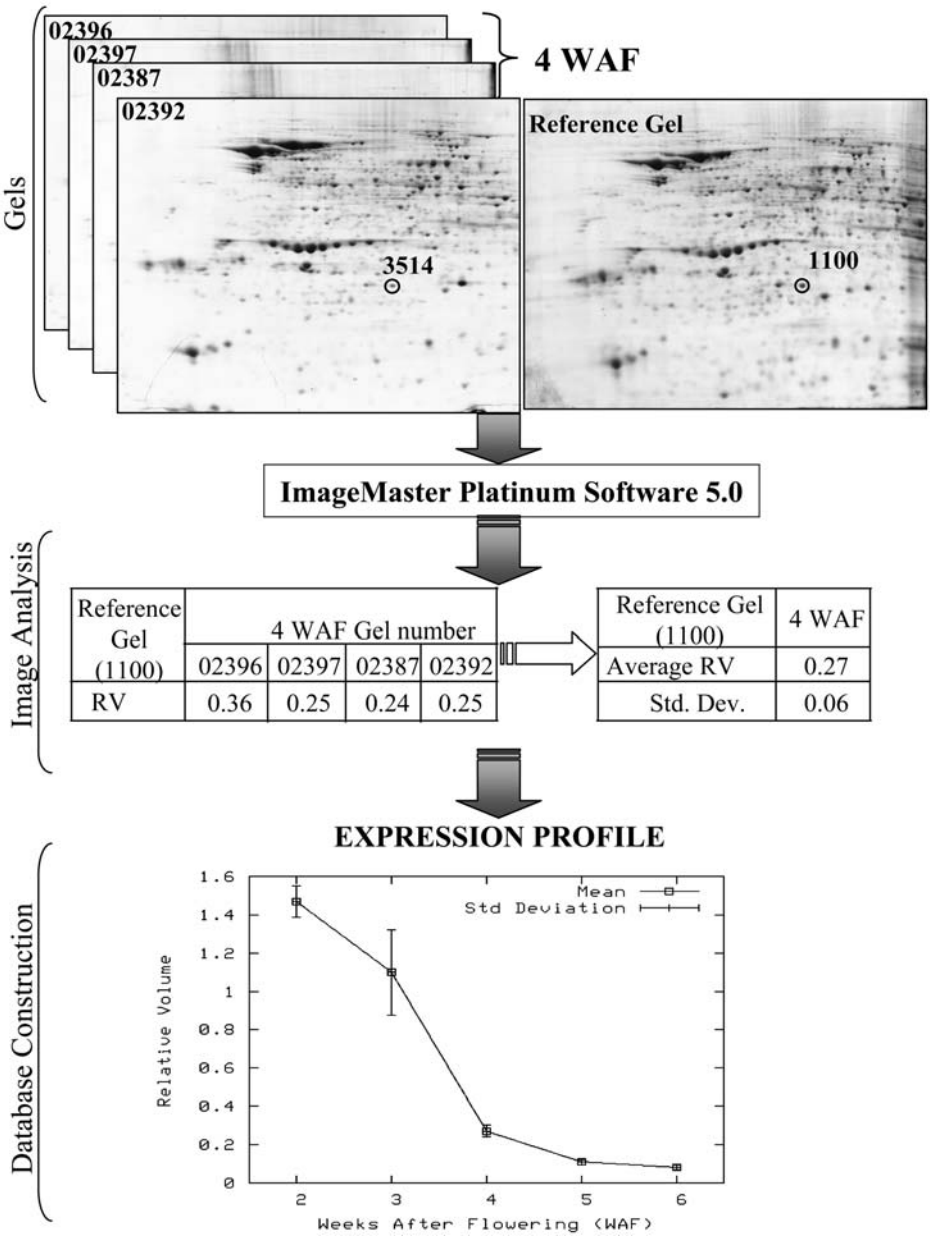

Figure 3. Experimental design for protein expression analysis. Four biological replicates for each developmental stage were analyzed using 2-DE. For quantification, only high-quality protein spots that were present in at least three out of the four gels were analyzed. Expression analyses were carried out using ImageMaster 2-D Platinum software version 5.0. Spots from analyzed gels were matched with spots from reference (pooled) gel and relative volumes (RV) were calculated for each spot group. Average relative volume for matched spots from replicate gels was calculated, and data were plotted onto a line graph. Finally expression profile data were uploaded onto a Web-based data repository (http://oilseedproteomics.missouri.edu/).

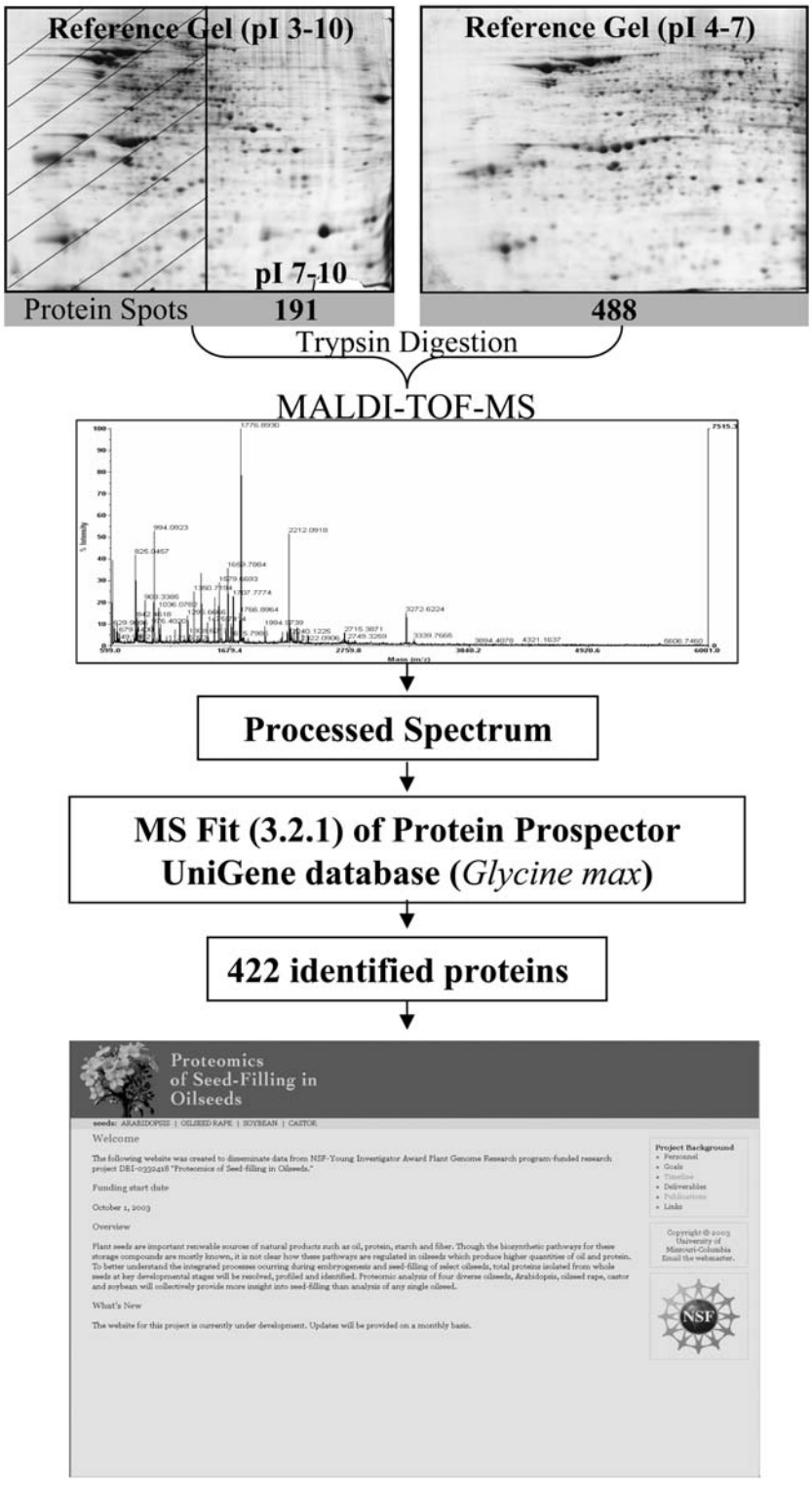

Figure 4. Experimental design for protein identification analyses using MALDI-TOF-MS. Protein spots with expression profiles were excised from reference (pooled) gels from pH ranges 3 to 10 and 4 to 7. For the pH 3 to 10 gel, spots were excised only from the pH 7 to 10 region to minimize redundancy. Each spot was trypsin digested, and peptides were analyses using MALDI-TOF-MS. Resultant spectra were searched against the NCBI soybean UniGene database using the MS-Fit program of Protein Prospector. All 422 identified proteins are listed in Table I and can also be accessed from a Web database.

time-of-flight mass spectrometry (MALDI-TOF-MS; Fig. 4). After MALDI-TOF analysis of tryptic peptides, mass tags were searched against the soybean UniGene database using the MS-FIT program of Protein Prospector (University of California, San Francisco, CA). The identification output was a list of 50 possible identifications (candidates) sorted according to their Molecular Weight Search (MOWSE) scores (Pappin et al., 1993). All candidates with MOWSE scores lower than 1×10^3 or with protein coverage less than 4% were discarded. Those with MOWSE scores greater than 1×10^6 were selected as positive assignments.

Candidates with MOWSE scores between 1×10^3 and 1×10^6 were subject to additional selection criteria. The theoretical molecular weight (MW) and pI should not differ more than 25% from the gel MW/pI value obtained from 2-DE gel calibration. If multiple candidates satisfied these criteria, the candidate with the highest MOWSE score was selected. Using this systematic approach, 422 proteins out of 679 (62%) were identified (Table I). One unique protein was often represented by more than one spot on the 2-DE gel, most likely due to posttranslation modifications or genetic isoforms. Taking into account this redundancy, 216 unique proteins out of 422 were identified.

Protein Classification and Assembly of a Web-Based Database

Each of the 422 identified proteins was classified into one of 15 functional categories as established by Bevan et al. (1998; Table I). The largest functional class, 82 proteins, was associated with metabolism. Within the metabolic protein class, a large subclass was formed by proteins associated with lipid and sterol metabolism (27). Not surprisingly, the second largest functional class was comprised of proteins associated with protein destination and storage, including 52 spots assigned to the seed storage proteins β-conglycinin and glycinin. Frequency distribution of the 216 unique proteins is shown in Figure 5. The largest group is represented by unclassified proteins (28%), followed by metabolic proteins (22%) and destination and storage proteins (10%).

A database, constructed and appended for archiving the data from this investigation, was the Web portal for proteomics research on oilseed plants (http://oilseedproteomics.missouri.edu/). This Web database offers two result viewing options: (1) a list of identified proteins sorted into plant functional categories with identification numbers linked to expression profile and identification data; and (2) interactive images of pooled gels pH 4 to 7 and pH 3 to 10 with active spot links that will lead users to expression and identification data. The protein identification data contain basic parameters describing PMF data and also a list of masses and intensities of tryptic peptides to allow for independent validation and updating of protein assignments.

DISCUSSION

Seed development can be simplified by categorizing this complex process into three sequential, temporal phases: embryogenesis, seed filling, and maturation. The seed-filling phase of development is the longest phase and is a period of cell division and cell expansion as well as storage product synthesis (Mienke et al., 1981). The metabolic events that occur during seed filling determine the overall storage composition of seed, and targeted transgenic alteration of these pathways can have a major impact on seed quality traits (Thelen and Ohlrogge, 2002). Despite its importance in quantitative trait determination, the seed-filling phase of oilseed development is poorly understood from a regulatory perspective. A recent investigation on the proteomics of seed filling in *M. truncatula* has been reported (Gallardo et al., 2003). The authors profiled and identified 84 proteins at various stages of *M. truncatula* seed development, and 23 of those identified proteins were storage proteins specific for Medicago. Interestingly, only 15 proteins from this study on soybean seed filling also were observed in Medicago, a related legume. Many of those common proteins, including Suc synthase and Met synthase, shared similar developmental expression profiles. Recently, two studies were published on proteomics of mature soybean seeds; Herman et al. (2003) identified 111 proteins, and Mooney and Thelen (2004) identified 44 proteins. When compared with this study, 55 and 22 proteins, respectively, were also identified and characterized during the seed-filling period of the current investigation.

Composite Expression Profiles of Plant Functional Classes Reveal Different Expression Trends

To characterize global expression trends of proteins involved in different processes, we established composite expression profiles by summing protein abundance, expressed as relative volume, for each protein in each functional class for the five seed stages (Figs. 6 and 7). Relative abundances of metabolic proteins decreased during the experimental period (Fig. 6A), suggesting metabolic activity curtails as seeds approach maturity. Interestingly, proteins involved in lipid and sterol metabolism decreased from 2 to 4 WAF, but after 4 WAF their abundance slightly increased. The protein destination and storage class of proteins increased during late seed filling, and this was due to the preponderance of seed storage proteins (Figs. 6B and 7). The transporter class of proteins, which includes the ubiquitous Suc-binding proteins, exhibited almost constant but slightly increased expression during the experimental period (Fig. 6C). Disease- and defense-related proteins were highly abundant at the early stage of seed filling; later their abundance decreased to about 50% and from 4 WAF was stable. Proteins involved in energy production increased in abundance during seed filling (Fig. 7), whereas cell growth and division proteins as well as signal transduction proteins each had decreasing expression profiles. Proteins involved in protein synthesis and secondary metabolism decreased in abundance during the experimental period.

Unknown Proteins Account for 22% of All Identified Proteins

Out of 422 identified proteins, 92 were annotated as unknown. The classification of unknown proteins

Table I. *Proteins identified by PMF from soybean immature seed reference maps*

Soybean proteins were classified according to protein functional categories described by Bevan et al. (1998). Proteins were identified by MALDI-TOF-MS analysis of tryptic peptides following searching against the NCBI soybean UniGene database. The putative protein identifications with MOWSE score 1×10^6 and higher were considered as positive identifications, and those with MOWSE score lower than 1×10^3 or with protein coverage less than 4% were discarded. Putative protein identifications with MOWSE score between these two values were considered as positive identifications only if the difference between theoretical and experimental MW/pI values was not greater than 25%. The expression profile for each protein listed in the table is available online (http://oilseedproteomics.missouri.edu/). The table includes spot number, accession number for UniGene entry, and MOWSE score. Number of matched tryptic peptides/percentage of protein coverage (Match/%Cov) and the theoretical and experimental MW/pI values also are indicated. PS, Only partial sequence available.

	Spot	Accession No.	MOWSE Score	Match/ %Cov	Theoretical MW/pI	Experimental MW/pI
01 Metabolism						
01.01 Amino acid						
Asn synthetase	438	Gma.197	3.30E + 03	5/7	67.4/6.2	70.5/5.6
Asn synthetase 1	485	Gma.12045	1.52E + 03	6/9	65.9/6.3	63.9/5.4
Asn synthetase 2	2245	Gma.15625	6.11E + 03	6/12	66.1/6.0	56.6/6.8
Cys synthase	984	Gma.631	4.00E + 05	10/36	34.6/5.5	38.2/5.7
Cytosolic Gln synthetase	813	Gma.3518	7.09E + 03	8/18	40.0/5.7	47.1/5.5
Cytosolic Gln synthetase	829	Gma.3518	1.72E + 07	8/37	39.0/5.5	46.2/5.4
Ferredoxin-dependent Glu synthase	293	Gma.1390	4.07E + 03	6/8	110.2/6.0	95.9/6.4
L-Asparaginase	835	Gma.1179	2.66E + 03	5/13	36.1/5.0	45.7/5.7
Met synthase	281	Gma.2874	8.49E + 12	20/34	96.6/6.3	85.7/6.9
Unknown; similar to DIAMINOPIMELATE EPIMERASE	2583	Gma.6655	2.40E + 03	4/19	28.0/6.3	33.3/6.7
Unknown; similar to DEOXYHYPUSINE SYNTHASE	1037	5606288	1.07E + 03	5/15	28.6/4.7	34.6/4.9
Unknown; similar to DEOXYHYPUSINE SYNTHASE	1129	5606288	1.53E + 03	5/13	28.6/4.7	28.8/5.4
01.02 Nitrogen and sulfur						
Adenosine 5′-phosphosulfate reductase	457	Gma.5648	2.53E + 03	8/11	52.6/6.0	67.2/5.2
Adenosine 5′-phosphosulfate reductase	635	Gma.5648	7.96E + 03	7/14	52.6/6.0	53.8/6.4
Adenosine 5′-phosphosulfate reductase	1340	Gma.5648	1.98E + 04	9/14	52.6/6.0	43.0/6.5
Asp aminotransferase cytosolic isozyme AAT2	2423	Gma.3139	6.16E + 09	15/42	52.9/8.2	45.2/8
Ferric leghemoglobin reductase-2 precursor	2438	Gma.10721	2.59E + 03	7/10	53.2/7.3	42.4/8.4
Hydroxyisourate hydrolase	2256	Gma.8450	4.24E + 03	7/5	63.8/6.1	55.2/7.3
Hydroxyisourate hydrolase	440	Gma.8450	2.96E + 03	8/6	63.8/6.1	70.5/5.4
Hydroxyisourate hydrolase	474	Gma.8450	1.11E + 03	6/4	63.8/6.1	64.3/6.3
Hydroxyisourate hydrolase	493	Gma.8450	6.89E + 03	8/11	63.8/6.1	63.7/4.8
Hydroxyisourate hydrolase	663	Gma.8450	1.72E + 05	11/12	63.8/6.1	53.2/5.7
Ni-binding urease accessory protein UreG	2621	Gma.685	3.08E + 03	5/24	27.2/8.8	31.1/8.1
Ni-binding urease accessory protein UreG	946	Gma.685	1.04E + 07	12/54	30.2/5.9	39.7/6.2
Ni-binding urease accessory protein UreG	942	Gma.685	3.37E + 06	11/45	39.9/6.2	29.2/5.8
Nitrate reductase	319	Gma.1221	4.31E + 03	7/4	99.8/6.3	90.1/6.5
Nitrate reductase	373	Gma.1221	4.73E + 03	7/6	99.8/6.3	81.6/5.6
Soybean late nodulin	830	Gma.8424	1.06E + 05	10/14	61.6/6.6	45.9/6.7
Nodulin 35	2411	Gma.2854	6.49E + 03	7/22	35.7/7.9	45.3/8.6
01.03 Nucleotides						
Nucleoside diphosphate kinase	1282	Gma.1449	5.39E + 07	10/61	17.7/7.1	18.5/6.5
Nucleoside diphosphate kinase	2881	Gma.1449	1.83E + 05	7/35	17.7/7.1	17.7/6.7
01.05 Sugars and polysaccharides						
α-Galactosidase	691	Gma.2570	3.17E + 03	8/13	46.4/6.1	52.3/4.9
α-Galactosidase	2427	Gma.2570	3.39E + 03	6/12	46.4/6.1	43.5/7.2
β-1,3-Endoglucanase	2486	Gma.2359	3.07E + 04	4/60	39.5/8.7	35.8/7.3
Endo-xyloglucan transferase	2577	Gma.2240	3.36E + 04	8/27	34.4/8.4	32.4/9.1
Enolase	557	Gma.17597	2.46E + 04	9/29	49.6/6.2	58.4/5.5

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

	Spot	Accession No.	MOWSE Score	Match/ %Cov	Theoretical MW/pI	Experimental MW/pI
Enolase	558	Gma.17597	4.15E + 03	6/45	49.6/6.2	58.6/5.6
GTP-binding protein	1147	Gma.1450	5.76E + 03	5/19	27.4/6.3	27.1/6.0
Mannosyl-oligosaccharide 1,2-α-mannosidase	520	Gma.5676	1.13E + 03	5/9	66.1/6.1	62.4/5.3
NADPH-specific isocitrate dehydrogenase	718	Gma.8359	9.77E + 10	21/47	49.1/6.1	50.7/6.0
NADPH-specific isocitrate dehydrogenase	2387	Gma.8359	1.00E + 03	5/12	49.1/6.1	48.3/6.8
Suc synthase	225	Gma.1828	1.16E + 09	18/21	92.3/6.1	108.2/6.1
Suc synthase	229	Gma.1828	6.70E + 14	26/32	92.3/6.1	106.5/6.3
Suc synthase	232	Gma.1828	2.63E + 24	39/50	92.3/6.1	106.2/6.4
Suc synthase	233	Gma.1828	1.07E + 17	28/40	92.3/6.1	106.5/6.4
Suc synthase	234	Gma.1828	1.02E + 25	45/51	92.3/6.1	105.8/6.3
Suc synthase	237	Gma.1828	4.46E + 22	37/49	92.3/6.1	104.1/6.5
UDP-Glc dehydrogenase	560	Gma.6350	7.22E + 04	9/22	53.4/6.1	59.5/5.8
Unknown; similar to AMYGDALIN HYDROLASE ISOFORM AH I PRECURSOR	2632	Gma S4901787	6.68E + 04	6/20	32.8/7.3	29.7/9.2
Unknown; similar NAD-DEPENDENT ISOCITRATE DEHYDROGENASE PRECURSOR	1015	Gma.7096	1.62E + 03	4/21	27.3/7.1	35.5/6.1
Unknown; similar to SW:UDP-*N*-ACETYLMURAMOYLALANYL-D-GLUTAMATE-2, 6-DIAMINOPIMELATE LIGASE	1237	Gma S5020437	3.55E + 03	5/25	18.6/4.6	20.9/5.4
Unknown; similar to TREHALOSE-6-PHOSPHATE SYNTHASE HOMOLOG	1026	Gma.5823	4.19E + 03	4/22	25.2/4.4	34.8/5.9
Unknown; similar to TREHALOSE-6-PHOSPHATE SYNTHASE HOMOLOG	1097	Gma.5823	5.67E + 03	4/22	25.2/4.4	31.0/5.8
Unknown; similar to TREHALOSE-6-PHOSPHATE SYNTHASE HOMOLOG	1123	Gma.5823	3.41E + 03	4/21	25.2/4.4	29.6/5.1
Unknown; similar to TREHALOSE-6-PHOSPHATE SYNTHASE HOMOLOG	1132	Gma.5823	5.67E + 03	4/22	25.2/4.4	28.7/5.4
01.06 Lipid and sterol						
α-Carboxyltransferase aCT-1 precursor	2146	Gma.8430	1.14E + 05	15/21	79.0/8.0	65.3/7.1
β-Ketoacyl-ACP synthetase 2	2148	Gma.5093	9.02E + 03	7/13	54.2/7.4	70.9/7.7
β-Ketoacyl-ACP synthetase I	672	Gma.248	4.32E + 04	9/17	52.4/8.1	52.7/6.6
Biotin carboxylase precursor	466	Gma.181	2.96E + 03	8/14	58.9/7.0	66.0/5.7
Biotin carboxylase precursor	541	Gma.181	7.30E + 03	7/12	58.9/7.0	59.2/5.9
Biotin carboxylase precursor	824	Gma.181	4.09E + 05	11/17	58.9/7.0	46.6/5.9
Biotin carboxylase precursor	483	Gma.8414	2.39E + 04	9/13	62.4/7.5	63.7/6.6
Lipoxygenase	2373	Gma.8458	3.01E + 03	5/14	40.5/5.8	48.3/7.2
Lipoxygenase	172	Gma.10969	3.39E + 04	9/5	96.8/6.0	114.0/6.4
Lipoxygenase	277	Gma.10969	1.02E + 04	8/4	96.8/6.0	96.3/6.7
Lipoxygenase	346	Gma.10969	2.06E + 04	8/5	96.8/6.0	87.6/5.4
Lipoxygenase	2081	Gma.10969	1.40E + 07	14/22	96.8/6.0	92.3/6.9
Lipoxygenase	2092	Gma.2019	3.81E + 03	7/8	96.3/6.3	89.5/7.1
Lipoxygenase	185	Gma.7625	1.26E + 07	12/16	97.7/6.0	114.3/6.0
Lipoxygenase	187	Gma.7625	4.96E + 11	19/26	97.7/6.0	112.9/6.0
Lipoxygenase	164	Gma.17610	3.68E + 03	7/4	95.8/5.8	115.7/6.1
Lipoxygenase	746	Gma.8458	1.61E + 04	6/17	40.5/5.8	49.5/5.3
Lipoxygenase	758	Gma.8458	1.79E + 03	4/12	40.5/5.8	49.5/5.1
Lipoxygenase	982	Gma.8458	1.27E + 03	4/14	40.5/5.8	37.6/6.4
Lipoxygenase	204	Gma.947	1.46E + 11	15/26	97.8/6.1	110.2/6.2
Lipoxygenase-2	206	Gma.1862	3.49E + 08	13/21	97.7/6.3	109.6/6.5

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

	Spot	Accession No.	MOWSE Score	Match/ %Cov	Theoretical MW/pI	Experimental MW/pI
Lipoxygenase-2	220	Gma.1862	3.86E + 18	30/40	97.7/6.3	107.2/6.6
Lipoxygenase-3	212	Gma.15534	2.24E + 16	33/40	97.6/6.2	108.2/6.6
Myoinositol-3-phosphate synthase	530	Gma.15564	9.96E + 05	13/37	56.8/5.4	61.3/5.4
Squalene synthase	771	Gma.8028	4.46E + 03	6/17	48.9/6.6	48.7/5.6
Unknown; similar to CER2 GENE	2526	Gma.3303	2.08E + 03	5/17	27.2/9.3	36.5/9.0
Unknown; similar to JASMONIC ACID 2	2726	Gma.5331	1.90E + 03	4/18	22.0/9.2	23.1/8.4
02 Energy						
02.01 Glycolysis						
Pyruvate kinase	2270	Gma.4073	1.12E + 09	13/27	57.5/7.7	54.5/7.7
Pyruvate kinase	2390	Gma.4073	1.53E + 04	7/12	57.5/7.7	46.9/7.3
02.02 Gluconeogenesis						
Glyoxysomal isocitrate lyase	648	Gma.52	2.87E + 03	8/10	62.8/7.8	53.7/6.3
Malate synthase	2194	Gma.8130	1.23E + 03	6/6	63.9/7.0	61.9/8.3
Malate synthase	2243	Gma.8130	3.41E + 03	7/10	63.9/7.0	58.1/6.9
Malate synthase	2314	Gma.8130	1.20E + 03	8/9	63.9/7.0	52.2/7.2
Malate synthase	381	Gma.8130	1.78E + 04	8/11	63.9/7.0	79.5/5.9
02.07 Pentose phosphate						
6-Phosphogluconate dehydrogenase	800	Gma.1689	7.90E + 03	7/13	58.4/5.8	47.6/5.5
6-Phosphogluconate dehydrogenase	2350	Gma.1689	2.16E + 04	8/15	58.4/5.8	50.3/7.1
2.10 TCA pathway						
NADP-dependent isocitrate dehydrogenase	2385	Gma.5543	3.28E + 03	8/10	52.4/9.0	47.3/7.4
2.13 Respiration						
Unknown; similar to CYTOCHROME *C*	2798	Gma.6660	2.15E + 03	4/30	15.6/9.6	20.6/8.6
Cytochrome P450 monooxygenase	934	Gma.140	2.58E + 05	11/17	PS	40.8/5.2
2.30 Photosynthesis						
Chlorophyll *a*/*b*-binding protein	1056	1053215	4.41E + 03	4/16	29.1/6.1	33.4/5.7
PEP carboxylase	163	Gma.62	2.16E + 04	10/5	111.0/6.2	117.2/5.7
03 Cell growth/division						
03.01 Cell growth						
β-Tubulin	535	Gma.5502	3.71E + 09	20/46	46.2/5.7	59.7/6.0
β-Tubulin	546	Gma.5502	2.14E + 06	16/41	46.2/5.7	58.2/6.0
β-Tubulin	591	Gma.5502	4.07E + 03	10/22	46.2/5.7	56.3/4.8
Syringolide-induced protein 19-1-5	929	Gma.117	2.96E + 03	7/12	32.0/5.3	41.7/4.8
Syringolide-induced protein 19-1-5	1110	Gma.117	3.02E + 03	6/10	32.0/5.3	30.1/6.0
Syringolide-induced protein 19-1-5	1164	Gma.117	3.02E + 03	6/10	32.0/5.3	25.8/5.9
Syringolide-induced protein 19-1-5	1176	Gma.117	1.02E + 04	7/10	32.0/5.3	24.5/6.1
03.16 DNA synthesis/replication						
DNA polymerase delta	106	Gma.158	5.00E + 05	13/5	127.4/8	147.3/6.3
03.22 Cell cycle						
Mitotic cyclin a2-type	2273	Gma.8421	1.93E + 03	7/19	55.5/8.6	55.6/7.0
Mitotic cyclin a2-type	2318	Gma.8421	1.28E + 03	6/15	55.5/8.6	50.7/8.7
03.25 Cytokinesis						
SDL (soybean dynamin-like protein)	2172	Gma.4720	9.32E + 03	8/13	68.9/7.7	65.1/7.2
03.26 Growth regulators						
GH1 protein	2414	Gma.3158	1.54E + 04	4/20	36.5/8.7	44.3/6.9
03.99 Other						
51-kD seed maturation protein	883	Gma.8467	7.85E + 03	8/12	51.6/6.7	43.5/6.1
51-kD seed maturation protein	2170	Gma.8467	5.76E + 03	7/12	51.6/6.7	63.3/7.8
51-kD seed maturation protein	2272	Gma.8467	3.21E + 03	6/9	51.6/6.7	55.8/8.5
51-kD seed maturation protein	2287	Gma.8467	9.64E + 03	8/12	51.6/6.7	53.9/6.8
Dehydrin-like protein	1105	Gma.2044	1.90E + 06	9/63	24.7/6.2	30.1/6.5
Lea protein	900	Gma.8434	5.89E + 03	6/10	50.1/7.3	42.5/6.4
Lea protein	2225	Gma.8434	5.47E + 03	7/14	50.1/7.3	59.9/7.0
Lea protein	2233	Gma.8434	6.14E + 03	8/13	50.1/7.3	56.8/8.1
Lea protein	2304	Gma.8434	5.72E + 03	6/13	50.1/7.3	53.0/8.1
Lea protein	2422	Gma.8434	2.53E + 03	6/12	50.1/7.3	45.8/6.8
Lea protein	2417	Gma.8434	1.49E + 03	7/10	50.1/7.3	44.9/8.4

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

	Spot	Accession No.	MOWSE Score	Match/ %Cov	Theoretical MW/pI	Experimental MW/pI
LEA protein	354	Gma.146	1.52E + 13	26/49	70.3/6.6	83.6/6.5
LEA protein	427	Gma.146	3.14E + 04	8/12	70.3/6.6	69.7/5.7
LEA protein	436	Gma.146	7.33E + 04	9/9	70.3/6.6	68.9/6.5
LEA protein	462	Gma.146	1.97E + 04	10/11	70.3/6.6	65.8/6.1
Maturation polypeptide	503	Gma.53	3.52E + 07	17/34	51.4/6.2	62.8/6.7
Maturation polypeptide	511	Gma.53	1.47E + 04	8/18	51.4/6.2	62.8/4.7
Maturation polypeptide	579	Gma.53	1.75E + 03	6/8	51.4/6.2	57.3/5.1
Maturation polypeptide	594	Gma.53	1.00E + 03	9/14	51.4/6.2	55.8/5.7
Maturation polypeptide	833	Gma.53	4.40E + 03	9/14	51.4/6.2	45.8/5.3
Maturation polypeptide	879	Gma.53	2.90E + 03	7/13	51.4/6.2	43.8/5.5
Seed maturation protein PM26	1089	Gma.8528	1.59E + 03	4/27	27.6/4.9	31.6/4.6
Seed maturation protein PM31	2721	Gma.8531	3.22E + 03	6/27	18.9/8.2	23.6/6.8
Seed maturation protein PM31	2786	Gma.8531	1.57E + 03	6/25	18.9/8.2	21.4/8.0
Seed maturation protein PM34	2537	Gma.2253	1.95E + 05	9/24	33.1/7.8	35.1/7.0
Seed maturation protein PM38	2780	Gma.8492	3.02E + 06	10/14	91.7/5.4	22.0/8.5
Seed maturation protein PM39	670	Gma.8491	1.28E + 03	6/10	46.7/5.7	53.1/5.4
Seed maturation protein PM39	836	Gma.8491	1.63E + 03	7/17	46.7/5.7	45.9/5.8
Seed maturation protein PM39	875	Gma.8491	2.38E + 04	8/11	46.7/5.7	43.9/4.9
Seed maturation protein PM39	897	Gma.8491	4.79E + 03	7/9	46.7/5.7	43.6/6.0
Seed maturation protein PM39	980	Gma.8491	3.23E + 03	6/13	46.6/5.7	38.5/5.1
Seed maturation protein PM39	1980	Gma.8491	6.12E + 07	13/31	46.6/5.7	144.5/6.8
Unknown; similar to TINY-LIKE PROTEIN [1]	2759	12492392	1.75E + 03	5/27	19.7/9.3	22.5/7.9
04 Transcription						
04.1901 General transcrition factors						
TGACG-motif binding factor	2439	Gma.8536	1.37E + 06	4/62	35.5/5.4	41.2/7.0
04.22 mRNA processing						
Unknown; similar to MAGO NASHI PROTEIN HOM	1220	Gma S5071218	4.63E + 05	10/35	23.9/8.5	21.9/6.4
05 Protein synthesis						
05.01 Ribosomal proteins						
Unknown; similar to SW:ARF_MAIZE P49076 ADP-RIBOSYLATION FACTOR	2834	Gma.6298	4.21E + 05	7/45	19.7/5.9	19.5/6.8
05.04 Translation factors						
Elongation factor 1-γ	616	Gma.322	1.31E + 04	8/14	48.3/6.3	54.7/6.0
Elongation factor 1-γ	636	Gma.322	5.54E + 03	6/12	48.3/6.3	53.7/6.2
Elongation factor 1-γ	840	Gma.322	1.99E + 03	6/12	48.3/6.3	45.5/5.6
Elongation factor 1-γ	880	Gma.322	2.26E + 03	6/13	48.3/6.3	43.4/6.2
Elongation factor 1-γ	932	Gma.322	4.61E + 03	8/13	48.3/6.3	40.9/5.2
Unknown; similar to SW:IF3A_TOBAC Q40554 EUKARYOTIC TRANSLATION INITIATION FACTOR 3 SUBUNIT 10	1156	Gma S5030297	3.40E + 03	6/21	26.6/6.4	26.3/5.8
Unknown; similar to SW:IF52_NICPL P24922 INITIATION FACTOR 5A-2	1246	Gma.15407	6.40E + 03	5/31	19.9/6.0	20.5/5.8
06 Protein destination and storage						
06.01 Folding and stability						
BiP isoform A	417	Gma.45	4.10E + 03	10/7	75.6/5.2	73.4/6.0
BiP isoform A	479	Gma.45	9.32E + 03	6/9	75.6/5.2	64.1/4.8
BiP isoform B	298	Gma.17631	4.43E + 06	11/23	73.9/5.1	94.8/5.4
BiP isoform B	357	Gma.17631	1.49E + 10	17/29	73.9/5.1	82.4/5.0
BiP isoform B	472	Gma.17631	1.71E + 04	6/16	73.9/5.1	64.7/6.3
BiP isoform B	473	Gma.17631	2.13E + 04	9/17	73.9/5.1	64.3/6.5
BiP isoform B	508	Gma.17631	1.75E + 06	10/18	73.9/5.1	62.2/5.6
BiP isoform B	528	Gma.17631	2.80E + 06	10/19	73.9/5.1	60.9/5.8
BiP isoform B	866	Gma.9527	1.15E + 06	10/19	73.8/5.1	44.3/5.2
BiP isoform B	941	Gma.9527	2.26E + 05	10/18	73.8/5.1	40.1/5.8

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

	Spot	Accession No.	MOWSE Score	Match/ %Cov	Theoretical MW/pI	Experimental MW/pI
BiP isoform B	2154	Gma.9527	1.08E + 06	8/14	73.8/5.1	69.3/7.9
BiP isoform B	2158	Gma.17631	4.53E + 03	7/8	73.9/5.1	70.9/6.7
06.13 Proteolysis						
20S proteasome subunit	1162	7839484	2.91E + 04	8/22	26.4/4.8	25.9/4.4
34-kD maturing seed vacuolar thiol protease precursor	722	Gma.1816	2.02E + 09	21/35	42.8/5.7	50.6/6.0
Cys proteinase	2471	Gma.434	7.61E + 03	5/14	42.1/5.8	36.9/7.3
Gag-protease polyprotein	2142	Gma.8464	7.90E + 03	8/7	70.4/7.9	67.6/7.8
Gag-protease polyprotein	583	Gma.8464	7.28E + 03	11/8	70.4/7.9	56.9/6.2
Kunitz trypsin inhibitor; KTi	1219	Gma.13030	8.62E + 05	9/33	26.6/5.8	22.1/4.8
Kunitz trypsin inhibitor; KTi	1236	Gma.13030	2.35E + 04	7/27	26.6/5.8	21.1/4.6
Kunitz trypsin inhibitor; KTi	1241	Gma.13030	2.35E + 04	7/27	26.6/5.8	20.8/4.5
Kunitz trypsin inhibitor; KTi	1247	Gma.13030	1.84E + 06	11/46	26.6/5.8	20.4/4.4
Subtilisin-type protease precursor	270	Gma.227	1.19E + 08	17/30	83.4/7.0	98.0/6.5
Subtilisin-type protease precursor	413	Gma.227	1.34E + 05	14/25	83.4/7.0	70.1/6.6
Unknown; similar to ENDOGENOUS α-AMYLASE/SUBTILISIN INHIBITOR	1182	Gma S5146395	1.76E + 03	4/16	24.9/5.1	24.0/5.0
Unknown; similar to MULTICATALYTIC ENDOPEPTIDASE COMPLEX	1192	Gma.16887	1.93E + 03	4/31	18.8/5.2	23.7/5.5
Unknown; similar to MULTICATALYTIC ENDOPEPTIDASE COMPLEX	1190	Gma.16887	8.05E + 03	5/37	18.8/5.2	23.7/5.3
06.20 Storage proteins						
β-Conglycinin α′-subunit	244	Gma.2270	2.31E + 15	28/41	74.5/5.8	104.8/5.6
β-Conglycinin α′-subunit	248	Gma.2270	1.14E + 11	19/32	74.5/5.8	103.8/5.5
β-Conglycinin α′-subunit	250	Gma.2270	4.33E + 08	17/29	74.5/5.8	103.8/5.6
β-Conglycinin α′-subunit	251	Gma.2270	1.83E + 06	11/16	74.5/5.8	102.4/5.4
β-Conglycinin α′-subunit	252	Gma.2270	4.16E + 05	11/17	74.5/5.8	102.4/5.4
β-Conglycinin α′-subunit	255	Gma.2270	3.94E + 16	27/41	74.5/5.8	99.7/5.3
β-Conglycinin α′-subunit	262	Gma.2270	4.25E + 10	20/27	74.5/5.8	97.7/5.3
β-Conglycinin α′-subunit	265	Gma.2270	1.38E + 19	33/50	74.5/5.8	98.4/5.2
β-Conglycinin α′-subunit	268	Gma.2270	4.72E + 15	30/50	74.5/5.8	95.6/5.1
β-Conglycinin α′-subunit	271	Gma.2270	9.20E + 10	18/33	74.5/5.8	98.7/4.9
β-Conglycinin α′-subunit	390	Gma.2270	4.47E + 04	10/16	74.5/5.8	77.9/5.1
β-Conglycinin α′-subunit	459	Gma.2270	6.38E + 05	10/18	74.5/5.8	65.6/5.2
β-Conglycinin α′-subunit	495	Gma.2270	2.72E + 09	18/27	74.5/5.7	63.9/5.1
β-Conglycinin α′-subunit	534	Gma.2270	2.30E + 03	9/5	74.5/5.8	59.0/6.3
β-Conglycinin α′-subunit	537	Gma.2270	1.79E + 09	16/25	74.5/5.7	60.7/5.1
β-Conglycinin α′-subunit	564	Gma.2270	4.87E + 09	18/26	74.5/5.8	59.0/5.0
β-Conglycinin α-subunit	316	Gma.17512	1.77E + 05	9/18	72.5/5.3	86.9/4.7
β-Conglycinin α-subunit	324	Gma.17512	4.99E + 09	16/25	72.5/5.3	90.1/4.5
β-Conglycinin α-subunit	326	Gma.17512	2.67E + 13	21/33	72.5/5.3	87.6/4.9
β-Conglycinin α-subunit	329	Gma.17512	3.87E + 06	11/19	72.5/5.3	88.0/5.0
β-Conglycinin α-subunit	342	Gma.17512	5.78E + 09	17/22	72.5/5.3	91.9/5.2
β-Conglycinin α-subunit	471	Gma.17512	1.88E + 09	17/26	72.5/5.3	65.2/4.9
β-Conglycinin α-subunit	523	Gma.17512	4.87E + 03	7/12	72.5/5.3	61.5/5.3
β-Conglycinin α-subunit	600	Gma.5834	8.13E + 07	14/23	70.3/5.1	55.2/5.5
β-Conglycinin α-subunit	638	Gma.1017	2.56E + 07	11/29	51.2/5.9	54.1/5.9
β-Conglycinin α-subunit	650	Gma.1017	2.24E + 13	21/50	51.2/5.9	53.7/6.0
Ferritin	2785	Gma.1089	1.64E + 03	5/14	28.2/5.7	22.1/7.0
Ferritin light chain	1170	Gma.18200	8.30E + 04	8/24	31.3/5.5	25.1/5.3
Glycinin A2B1a precursor	509	Gma.1857	6.83E + 03	6/14	54.8/5.6	61.8/5.7
Glycinin A2B1a precursor	596	Gma.1857	7.94E + 02	4/8	54.8/5.6	56.0/4.4
Glycinin A2B1a precursor	881	Gma.1857	4.02E + 03	5/11	54.8/5.6	43.7/5.4
Glycinin A2B1a precursor	950	Gma.1857	2.15E + 08	13/30	54.4/5.5	39.6/4.9
Glycinin A2B1a precursor	964	Gma.1857	3.18E + 08	14/31	54.8/5.6	38.2/5.2
Glycinin A2B1a precursor	952	Gma.1857	6.16E + 06	9/24	54.4/5.5	39.8/4.8

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

	Spot	Accession No.	MOWSE Score	Match/ %Cov	Theoretical MW/pI	Experimental MW/pI
Glycinin A2B1a precursor	953	Gma.1857	2.15E + 08	13/30	54.4/5.5	39.6/4.9
Glycinin A2B1a precursor	969	Gma.1857	9.90E + 06	13/36	54.4/5.5	38.5/5.1
Glycinin A2B1a precursor	2551	Gma.1857	6.39E + 07	11/22	54.4/5.5	34.3/7.2
Glycinin A2B1a precursor	2555	Gma.1857	2.47E + 07	11/24	54.4/5.5	34.4/7.9
Glycinin A2B1a precursor	2559	Gma.1857	1.59E + 08	12/27	54.4/5.5	34.3/7.3
Glycinin A2B1a precursor	2792	Gma.1857	1.74E + 05	8/17	54.4/5.5	21.0/9.0
Glycinin A5A4B3 subunits precursor	944	Gma.1812	1.38E + 06	14/20	30.9/6.5	39.8/4.5
Glycinin A5A4B3 subunits precursor	945	Gma.1812	6.58E + 07	18/25	30.9/6.5	39.8/4.6
Glycinin precursor	492	Gma.1897	9.19E + 06	10/22	57.5/6.6	63.5/6.1
Glycinin precursor	497	Gma.1897	6.15E + 09	17/40	57.5/6.6	62.6/6.2
Glycinin precursor	921	Gma.1897	1.03E + 08	12/30	58.0/5.6	41.3/5.6
Glycinin precursor	930	Gma.1897	3.41E + 09	16/42	58.0/5.6	40.6/5.5
Glycinin precursor	936	Gma.1897	4.71E + 08	14/36	58.0/5.6	39.4/5.4
Glycinin precursor	955	Gma.1897	4.42E + 08	12/29	58.0/5.6	38.7/5.3
Glycinin precursor	1053	Gma.1897	1.24E + 08	11/24	55.7/5.9	33.4/5.5
Glycinin	787	Gma.8205	1.08E + 03	5/12	58.4/5.5	48.2/6.0
Glycinin	845	Gma.8205	1.98E + 07	13/26	58.2/5.5	44.8/6.6
Glycinin	847	Gma.8205	8.48E + 05	9/25	58.2/5.5	44.9/6.3
Glycinin	2269	Gma.8205	2.36E + 07	10/23	58.2/5.5	55.9/8.4
Glycinin	2424	Gma.8205	2.19E + 07	12/26	58.2/5.5	42.9/6.8
07 Transporters						
07.01 Ions						
Calnexin	482	Gma.6427	2.28E + 04	10/16	63.2/4.8	64.5/5.8
Calnexin	700	Gma.6427	8.66E + 03	9/13	63.2/4.8	51.9/5.7
Calnexin	786	Gma.6427	6.35E + 05	10/13	63.2/4.8	48.0/4.2
Phosphate transporter	2456	Gma.292	3.93E + 03	4/19	38.7/9.2	41.6/8.2
Plasma membrane Ca^{2+}-ATPase	290	Gma.1047	2.09E + 04	10/9	111.0/5.7	96.6/5.9
Plasma membrane Ca^{2+}-ATPase	146	Gma.1178	1.25E + 03	6/7	111.4/5.9	129.2/5.0
Putative cytochrome P450	2163	Gma.155	5.34E + 03	7/11	57.7/8.3	66.8/6.9
07.07 Sugars						
Suc-binding protein	372	Gma.1872	1.66E + 06	10/16	61.6/6.3	77.1/6.2
Suc-binding protein	376	Gma.1872	1.54E + 06	11/17	61.6/6.3	78.3/6.1
Suc-binding protein	383	Gma.1872	1.21E + 11	23/38	61.6/6.3	75.0/6.3
Suc-binding protein	386	Gma.1872	3.45E + 07	14/19	61.6/6.3	79.1/6.0
Suc-binding protein	405	Gma.1872	2.86E + 04	8/13	61.6/6.3	77.1/5.8
Suc-binding protein	463	Gma.1872	4.10E + 03	7/12	61.6/6.3	65.6/6.2
Suc-binding protein	478	Gma.1872	3.72E + 05	10/16	61.6/6.3	63.7/6.6
Suc-binding protein	515	Gma.1872	2.92E + 06	12/21	61.6/6.3	61.8/6.0
Suc-binding protein	522	Gma.1872	2.02E + 06	11/19	61.6/6.3	61.3/6.1
Suc-binding protein	631	Gma.1872	1.25E + 05	8/17	61.6/6.3	54.2/5.8
Suc-binding protein	704	Gma.1872	2.87E + 03	7/9	61.6/6.3	51.8/5.1
Suc-binding protein	754	Gma.1872	1.89E + 04	7/15	61.6/6.3	49.4/5.2
Suc-binding protein	2186	Gma.1872	8.80E + 04	10/15	61.6/6.3	61.5/6.9
Suc-binding protein homolog S-64	2199	Gma.8486	2.97E + 03	5/10	56.4/6.2	60.7/7.9
Suc-binding protein homolog S-64	2303	Gma.8486	7.81E + 04	9/21	56.4/6.2	53.5/7.0
Suc-binding protein homolog S-64	2301	Gma.8486	2.48E + 03	7/8	56.4/6.2	52.4/8.0
Suc-binding protein homolog S-64	2311	Gma.8486	4.64E + 05	9/21	56.4/6.2	52.3/7.3
Suc-binding protein homolog S-64	2295	Gma.8486	1.19E + 04	7/13	56.4/6.2	51.8/7.8
Suc-binding protein homolog S-64	428	Gma.8486	5.70E + 10	20/35	56.4/6.2	70.1/6.1
Suc-binding protein homolog S-64	444	Gma.8486	4.24E + 04	9/18	56.4/6.2	69.3/6.0
Suc-binding protein homolog S-64	445	Gma.8486	1.17E + 06	9/15	56.4/6.2	68.5/6.4
Suc-binding protein homolog S-64	599	Gma.8486	2.98E + 03	6/9	56.4/6.2	55.2/6.3
Suc-binding protein homolog S-64	610	Gma.8486	2.26E + 03	6/15	56.4/6.2	55.2/5.1
07.25 ABC-type						
ABC transporter-like protein	399	Gma.4290	5.59E + 03	8/13	72.3/6.9	77.5/5.7
ABC transporter-like protein	416	Gma.4290	6.44E + 03	7/10	72.3/6.9	73.8/5.9
ABC transporter-like protein	480	Gma.4290	3.51E + 05	10/16	72.3/6.9	64.7/5.9
ABC transporter-like protein	496	Gma.4290	5.13E + 03	6/11	72.3/6.9	63.2/5.5
ABC transporter-like protein	555	Gma.4290	7.14E + 04	11/13	72.3/6.9	59.5/5.7
ABC transporter-like protein	571	Gma.4290	1.30E + 05	11/17	71.3/6.4	58.0/4.8

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

	Spot	Accession No.	MOWSE Score	Match/ %Cov	Theoretical MW/pI	Experimental MW/pI
ABC transporter-like protein	809	Gma.4290	1.30E + 05	11/14	71.3/6.4	49.4/6.1
ABC transporter-like protein	816	Gma.4290	1.72E + 05	10/15	71.3/6.4	50.3/5.6
ABC transporter-like protein	2091	Gma.4290	9.12E + 03	8/10	72.3/6.9	90.5/7.0
ABC transporter-like protein	2161	Gma.4290	5.55E + 03	10/16	72.3/6.9	71.7/7.0
ABC transporter-like protein	2191	Gma.4290	1.43E + 03	6/10	72.3/6.9	61.5/8.5
ABC transporter-like protein	2195	Gma.4290	2.19E + 05	12/17	72.3/6.9	61.1/8.1
ABC transporter-like protein	2230	Gma.4290	2.43E + 07	12/24	72.3/6.9	56.8/8.0
ABC transporter-like protein	2122	Gma.4290	6.89E + 05	11/19	72.3/6.9	82.4/8.1
ABC transporter-like protein	2205	Gma.4290	6.07E + 04	9/15	72.3/6.9	61.4/6.8
08 Intracellular traffic						
08.04 Mitochondrial						
Unknown; similar to ALANYL-TRNA SYNTHETASE	2594	Gma S4877143	5.70E + 03	6/15	31.0/8.1	31.5/9.3
Unknown; similar to ALANYL-TRNA SYNTHETASE	981	Gma S4877143	1.07E + 03	4/12	31.0/8.1	38.0/6.5
08.99 Others						
ADP-ribosylation factor	1245	Gma.8407	1.70E + 09	12/58	20.4/6.4	20.6/6.6
09 Cell structure						
09.01 Cell wall						
Unknown; similar to POLYGALACTURONASE ISOENZYME 1	2523	Gma.8368	7.21E + 03	8/18	27.7/9.6	34.8/9.2
Unknown; similar to XYLOGLUCAN ENDOTRANS-GLYCOSYLASE-RELATED PROTEIN	951	Gma S4904704	3.01E + 03	6/11	32.0/6.8	40.0/6.4
09.13 Chromosomes						
Cys-rich polycomb-like protein	166	Gma.8470	8.14E + 03	9/10	98.4/5.7	115.7/6.3
Cys-rich polycomb-like protein	167	Gma.8470	2.64E + 04	11/7	98.4/5.7	114.6/6.4
09.16 Mitochondria						
Unknown; similar to 36-KD OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN	2654	Gma.14833	6.83E + 05	9/37	27.0/9.2	29.9/7.9
Unknown; similar to 34-KD OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN	2560	Gma.10895	1.41E + 03	5/21	27.0/9.4	34.5/8.7
10 Signal transduction						
10.0404 Kinases						
Calmodulin-like domain protein kinase isoenzyme beta	574	Gma.8417	4.94E + 04	9/12	57.3/5.4	57.1/5.2
Calmodulin-like domain protein kinase isoenzyme beta	689	Gma.8417	1.67E + 03	6/11	57.3/5.4	52.2/5.8
GmCK1p	838	Gma.14820	2.51E + 03	6/13	41.2/5.7	45.4/6.0
GmCK2p	782	Gma.149	1.05E + 03	4/11	46.3/5.7	48.3/5.4
Protein kinase	2325	Gma.8418	1.28E + 04	9/22	41.0/6.1	52.2/6.7
Protein kinase	2310	Gma.1701	4.24E + 03	6/13	53.4/9.4	52.6/7.5
Protein kinase	818	Gma.8418	5.08E + 03	7/12	41.0/6.1	46.7/5.7
Protein kinase 3	2487	Gma.8419	1.23E + 04	7/16	41.8/5.7	36.8/7.2
Receptor-like protein kinase 2	1112	Gma.16818	5.27E + 04	5/29	25.3/4.8	29.7/6.3
Unknown; similar to NUCLEOSIDE DIPHOSPHATE KIN3	2863	Gma.2144	3.61E + 03	5/27	23.1/9.2	18.4/8.6
Unknown; similar to PROTEIN KINASE	2551	7147438	2.95E + 03	4/16	30.1/6.7	34.3/7.2
Unknown; similar to SERINE/THREONINE KINASE-LIKE PROTEIN	2799	Gma.17557	4.47E + 03	5/20	19.4/6.8	21.0/7.1
Unknown; similar to SERINE/THREONINE KINASE-LIKE PROTEIN	1263	Gma.17557	1.29E + 03	4/22	19.4/6.8	19.4/6.0
10.0410 G proteins						
G β-like protein	2580	Gma.3893	4.82E + 12	18/54	35.7/7.8	33.0/7.9

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

	Spot	Accession No.	MOWSE Score	Match/ %Cov	Theoretical MW/pI	Experimental MW/pI
G protein α-subunit	805	Gma.8047	4.81E + 03	8/21	48.2/5.6	47.4/5.8
G protein α-subunit	810	Gma.8047	6.63E + 03	7/22	48.2/5.6	47.3/5.7
10.99 Others						
Phosphoinositide-specific phospholipase C P12	442	Gma.8433	3.13E + 03	7/15	65.5/6.1	69.3/5.7
Phosphoinositide-specific phospholipase C P12	2188	Gma.8433	1.51E + 03	5/9	65.5/6.1	63.5/7.7
Phosphoinositide-specific phospholipase C P12	2247	Gma.8433	3.42E + 04	8/18	65.6/6.1	57.9/6.8
Unknown; similar to annexin	2579	Gma.5357	1.59E + 10	16/58	PS	34.2/7.6
11 Disease/defense						
11.01 Resistance genes						
Resistance protein LM12	684	Gma.8480	2.89E + 03	6/17	49.7/6.1	52.4/6.6
Resistance protein MG23	2340	Gma.8479	5.24E + 03	7/15	49.3/6.2	52.4/7.0
Resistance protein MG23	856	Gma.8479	2.03E + 03	6/8	49.3/6.2	44.8/5.2
11.05 Stress responses						
Heat shock protein	209	Gma.48	7.06E + 04	11/12	104.6/5.8	111.6/5.4
Sali3-2	937	Gma.6633	1.63E + 03	6/18	31.6/6.2	40.1/6.5
sti (stress inducible protein)	576	Gma.8435	1.84E + 04	9/18	64.3/5.8	57.3/5.3
sti (stress inducible protein)	2208	Gma.8435	1.49E + 03	6/8	64.3/5.8	60.4/7.1
11.06 Detoxification						
2,4-D inducible glutathione *S*-transferase	2623	Gma.339	1.48E + 03	6/16	25.6/6.2	31.0/7.9
Alcohol dehydrogenase 1	2446	Gma.244	1.46E + 04	11/38	40.1/6.2	41.2/7.0
Alcohol dehydrogenase 1	750	Gma.244	2.33E + 03	4/9	40.0/6.2	49.7/6.2
Alcohol dehydrogenase 1	763	Gma.244	8.99E + 03	8/17	40.0/6.2	48.6/6.5
Alcohol dehydrogenase 1	927	Gma.244	3.22E + 03	9/15	40.0/6.2	41.1/6.2
Ascorbate peroxidase	1095	310560	6.50E + 08	17/58	27.1/5.5	31.1/5.7
Ascorbate peroxidase 2 (APx2)	1100	Gma.1246	1.67E + 11	17/75	27.2/5.7	30.8/5.9
Catalase	2297	Gma.587	1.65E + 13	24/48	58.3/7.5	53.8/7.4
Catalase	2332	Gma.587	1.20E + 03	7/11	58.3/7.5	49.0/8.6
Catalase	2389	Gma.587	4.12E + 03	6/13	58.3/7.5	47.3/7.7
Catalase	568	Gma.2444	2.26E + 04	8/17	58.0/6.9	57.3/6.6
Catalase	611	Gma.2444	1.06E + 05	10/28	58.0/6.9	55.4/5.9
Catalase	717	Gma.2444	1.07E + 04	8/17	58.0/6.9	50.8/6.1
Catalase	418	Gma.587	4.25E + 03	6/11	58.3/7.5	73.4/6.7
Catalase	812	Gma.587	3.64E + 04	8/11	58.3/7.5	47.1/6.3
δ-Aminolevulinic acid dehydratase	766	Gma.6038	2.75E + 04	8/29	45.6/6.8	49.0/5.6
Glutathione *S*-transferase GST 23	2618	Gma.428	1.02E + 03	5/13	28.9/9.7	31.0/8.3
Glutathione *S*-transferase GST 23	2712	Gma.428	6.14E + 03	6/17	28.9/9.7	23.7/8.1
Glutathione *S*-transferase GST 23	2763	Gma.428	4.97E + 03	6/17	28.9/9.7	22.4/8.3
Glutathione *S*-transferase GST 7	2713	Gma.8518	4.97E + 03	6/26	27.6/6.5	23.7/8.3
12 Unclear classification						
Unknown; similar to DNA-BINDING PROTEIN	2631	Gma.9522	3.24E + 03	7/27	24.0/9.2	30.8/7.8
Unknown; similar to DNA-BINDING PROTEIN	2714	Gma.9522	1.07E + 04	6/32	24.0/9.2	22.9/9.2
Unknown; similar to NUCLEIC ACID-BINDING PROTEIN	1068	Gma.6247	9.24E + 03	5/34	26.9/6.4	32.7/6.6
Unknown; similar to NUCLEIC ACID-BINDING PROTEIN	1184	Gma.6247	1.94E + 03	5/25	26.9/6.4	23.9/6.0
Unknown; similar to NUCLEIC ACID-BINDING PROTEIN	1222	Gma.6247	1.21E + 03	4/28	26.9/6.4	21.9/6.7
Unknown; similar to PUTATIVE RNA-BINDING PROTEIN	1163	Gma.1793	2.15E + 03		31.2/4.5	25.8/4.8
13 Unclassified						
FAS2	2198	Gma.8476	3.14E + 03	6/10	64.2/6.2	61.5/7.5
FAS2	2367	Gma.8476	1.12E + 03	4/12	64.2/6.2	50.3/6.8
FAS2	655	Gma.8476	1.02E + 05	7/23	64.2/6.2	53.2/6.3
14-3-3-like protein	1046	Gma.596	1.88E + 03	5/25	33.8/5.0	33.9/6.2

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

	Spot	Accession No.	MOWSE Score	Match/ %Cov	Theoretical MW/pI	Experimental MW/pI
14-3-3 protein	1027	Gma.3665	1.95E + 06	4/24	24.9/4.9	34.5/4.7
14-3-3-like protein	1029	Gma.596	3.04E + 04	8/28	33.8/5.0	34.3/4.6
14-3-3-like protein	1081	Gma.3665	1.72E + 07	14/50	24.9/4.9	33.3/4.3
31-kD protein	2678	Gma.1822	3.96E + 06	4/80	28.9/8.6	26.0/8.3
Allergen Gly m Bd 28 K	2779	Gma.7309	1.33E + 08	8/20	52.6/5.7	22.0/7.1
CYP78A3p	2391	Gma.138	7.19E + 03	6/12	59.2/8.1	47.2/7.6
CYP83D1p	640	Gma.140	2.44E + 03	8/9	58.6/6.8	53.7/5.9
CYP83D1p	720	Gma.140	4.62E + 04	8/12	58.6/6.8	50.8/5.9
Seed coat BURP domain protein 1	2561	Gma.8496	8.92E + 03	8/15	34.5/7.8	35.2/7.7
SGF14A	1087	Gma.1177	1.75E + 04	9/27	29.2/4.7	31.7/4.4
SGF14A	1069	Gma.1177	1.88E + 04	10/40	29.2/4.7	32.9/4.4
SGF14B	1111	Gma.1417	1.16E + 03	4/20	27.9/4.8	30.1/4.6
SGF14C	1067	Gma.105	1.80E + 06	12/49	29.5/4.8	32.9/4.6
SGF14D	1036	Gma.1088	5.52E + 05	13/38	30.0/4.8	34.6/4.5
Valosin-containing protein	181	Gma.4013	2.33E + 09	18/20	89.8/5.2	114.3/5.1
Valosin-containing protein	171	Gma.4013	1.44E + 12	22/27	89.8/5.2	115.7/5.2
Unknown	2624	6073016	4.02E + 03	6/26	24.0/10.7	30.8/8.7
Unknown	849	5753418	2.02E + 03	7/27	34.2/6.4	44.9/6.7
Unknown	2585	Gma.976	1.46E + 04	5/17	32.2/9.2	33.6/7.3
Unknown	1251	10253941	1.40E + 04	6/36	22.0/5.0	20.4/5.6
Unknown	2832	Gma.7327	2.65E + 03	4/26	17.0/9.6	19.7/8.9
Unknown	2747	Gma.4645	5.01E + 03	5/38	20.2/5.4	23.1/7.1
Unknown	2827	Gma.4645	1.93E + 03	4/32	20.2/5.4	20.4/6.8
Unknown	2677	6073016	2.20E + 03	7/21	23.8/10.0	25.9/8.1
Unknown	2679	Gma.8075	7.33E + 03	6/21	21.5/7.0	26.5/8.7
Unknown	2692	Gma.63	5.91E + 03	8/27	24.6/9.8	26.4/9.1
Unknown	2760	Gma.63	8.18E + 03	8/27	24.6/9.8	22.5/8.2
Unknown	2809	Gma.63	7.03E + 03	9/35	24.6/9.8	20.5/8.1
Unknown; similar to HYPOTHETIC 12.9-KD PROTEIN	2806	Gma.515	2.47E + 03	4/25	21.1/5.4	21.3/6.9
Unknown; similar to HYPOTHETIC 55.8-KD PROTEIN	2729	Gma S4866178	1.32E + 03	5/21	26.4/5.8	23.5/7.3
Unknown; similar to HYPOTHETIC 28.7-KD PROTEIN	2660	Gma.2875	3.25E + 04	4/29	23.6/6.2	30.6/6.9
Unknown	2838	Gma.7120	1.55E + 04	7/26	19.7/10.3	19.6/8.4
Unknown; similar DNA-BINDING PROTEIN	2754	Gma S5106326	8.91E + 03	6/26	23.2/6.8	23.0/6.9
Unknown; similar to TR:Q40478 Q40478 EREBP-4	2627	5342373	1.54E + 03	5/17	30.4/9.1	31.1/8.5
Unknown; similar HYPOTHETIC 20.7-KD PROTEIN	2625	Gma.1247	1.12E + 03	5/9	24.2/9.2	31.0/7.6
Unknown; similar to 20.7 PROTEIN	2817	4259056	2.47E + 03	4/22	25.1/9.3	20.5/7.8
Unknown; similar to HYPOTHETIC 67.9-KD PROTEIN	1205	Gma.3403	4.81E + 03	6/25	18.0/6.1	22.9/5.9
Unknown; similar to HYPOTHETIC 67.9-KD PROTEIN	1227	Gma.3403	2.39E + 03	6/18	18.0/6.1	21.6/6.1
Unnumbered protein product	572	Gma.595	4.32E + 09	17/34	57.3/5.8	58.4/5.4
Unknown	1187	Gma.3293	2.63E + 03	5/25	20.4/5.0	23.7/5.6
RecA/Rad51/DMC1-like protein	874	Gma.8429	1.69E + 03	6/14	39.9/5.7	44.0/6.0
Unknown	1243	5770674	4.05E + 04	5/25	21.2/7.9	20.7/6.3
Unknown	2843	Gma.5464	1.28E + 03	4/25	16.1/9.2	19.5/7.6
Unknown	1109	Gma.8021	7.97E + 03	6/19	23.3/5.3	30.1/4.9
Unknown	957	12493322	3.09E + 03	4/18	30.2/6.4	40.3/6.7
Unknown	993	12493322	3.41E + 03	5/22	30.2/6.4	36.7/6.3
Unknown	1098	12493322	2.73E + 03	5/23	30.2/6.4	31.0/5.6
Unknown	1189	12492588	1.50E + 03	4/17	27.1/4.9	23.7/5.5
Unknown	2802	Gma.4645	1.39E + 04	6/45	20.2/5.4	21.0/6.6
Unknown	1107	16343862	2.53E + 03	4/21	27.5/7.1	30.1/6.7
Unknown	1229	Gma.4965	8.96E + 03	4/26	19.9/4.7	21.7/4.3
Unknown	1277	Gma.5233	3.19E + 03	5/23	22.2/5.4	18.8/5.9

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

	Spot	Accession No.	MOWSE Score	Match/ %Cov	Theoretical MW/pI	Experimental MW/pI
Unknown	1275	Gma.8268	1.10E + 04	6/36	16.0/4.8	18.7/5.5
Unknown	1137	26056530	1.02E + 03	7/15	30.7/5.6	28.6/5.8
Unknown	1117	26058275	3.81E + 03	6/14	27.6/6.3	29.8/5.3
Unknown	2541	Gma.4664	2.58E + 03	6/19	29.8/7.5	35.1/6.8
Unknown	2582	Gma.69	1.20E + 06	5/41	PS	33.9/7.6
Unknown	2723	Gma.8139	6.48E + 03	6/23	23.1/7.1	23.3/7.5
Unknown; similar to HYPOTHETIC 37.6-KD PROTEIN	2619	Gma S4922039	1.82E + 03	5/21	27.7/6.6	32.0/6.8
Unknown; similar to HYPOTHETIC 68.6-KD PROTEIN	2647	Gma.6744	7.45E + 03	4/22	31.5/8.8	30.4/7.5
Unknown; similar to HYPOTHETIC 68.2-KD PROTEIN	2715	Gma S4891477	2.85E + 03	5/14	24.7/9.0	23.9/8.7
Unknown	2691	Gma.6780	1.01E + 03	6/17	28.4/7.0	25.9/7.2
Unknown	2656	Gma.11253	1.01E + 03	5/12	26.7/9.1	29.7/9.1
Unknown	2574	Gma.8719	4.18E + 04	6/18	27.5/8.6	34.0/8.1
Unknown; similar to HUMAN TAFAZZIN	1145	12488022	2.22E + 03	5/17	27.0/5.8	27.5/6.6
Unknown; similar HYPOTHETIC 32.6-KD PROTEIN	1016	Gma.4424	3.54E + 04	9/28	28.9/6.7	35.4/5.7
Unknown; similar to HOMEODOMAIN-CONTAINING PROTEIN	1261	Gma.9506	2.44E + 03	5/16	16.9/6.3	19.6/5.3
Unknown; similar to CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES ZINC-BINDING REGION	1186	Gma S4944655	5.05E + 03	5/24	23.0/6.2	23.8/6.7
20 Secondary metabolism						
20.1 Phenylpropanoids/phenolics						
4-Coumarate:CoA ligase isoenzyme 2	2192	Gma.4928	8.27E + 04	8/15	62.6/7.1	61.4/7.2
4-Coumarate:CoA ligase isoenzyme 2	2242	Gma.4928	3.07E + 05	10/16	62.6/7.1	55.3/8.8
4-Coumarate:CoA ligase isoenzyme 2	2312	Gma.4928	1.33E + 03	6/10	62.6/7.1	51.9/8.4
4-Coumarate:CoA ligase isoenzyme 2	377	Gma.4928	1.88E + 04	9/16	62.6/7.1	79.1/6.6
20.3 Alkaloids						
Unknown; similar to PUTATIVE STRICTOSIDINE SYNTHASE	1158	Gma.1660	1.06E + 03	5/14	32.5/5.3	26.0/5.5
20.99 Others						
Dihydroflavonol-4-reductase DFR1	2453	Gma.5313	1.25E + 03	4/12	39.2/8.3	43.6/7.8
Dihydroflavonol-4-reductase DFR1	2448	Gma.5313	1.40E + 03	4/12	39.2/8.3	41.9/8.6
Dihydroflavonol-4-reductase DFR1	2454	Gma.5313	5.73E + 05	9/22	39.2/8.3	41.0/8.0
Isoflavone reductase homolog 2	912	Gma.1950	2.31E + 08	11/43	34.1/5.5	41.7/5.7
Isoflavone reductase homolog 2	920	Gma.1950	3.95E + 06	10/30	34.1/5.5	41.6/5.9
Phytoene desaturase	661	Gma.51	3.46E + 05	9/16	63.6/7.1	48.6/6.5
γ-Glutamyl hydrolase	2573	Gma.1314	2.45E + 09	14/35	37.7/6.1	34.0/8.9

based on their expression profiles may provide clues about possible functions in seed development (Fig. 8). Based upon the similarities between expression profiles of unknown proteins and expression profiles for each major functional class (Figs. 6 and 7), it is possible to speculate on the function of these unknown proteins. Often expression of different isoforms of the same protein have different trends (i.e. Suc-binding proteins discussed in the next paragraph; 7 increased, 5 decreased). For this reason, the composite expression profile of the functional class taken for comparison needs to contain a significant number of proteins. Taking into account the number of identified proteins in this study, functional classes with 40 and more proteins were considered suitable for such analysis.

The largest group of unknown proteins (57%) decreased in expression during seed development. This expression profile is characteristic for proteins involved in metabolism (Fig. 6A) and cell growth/division (Fig. 7). Thirteen percent of unknown proteins increased in abundance during seed filling, and this was characteristic for proteins associated with protein

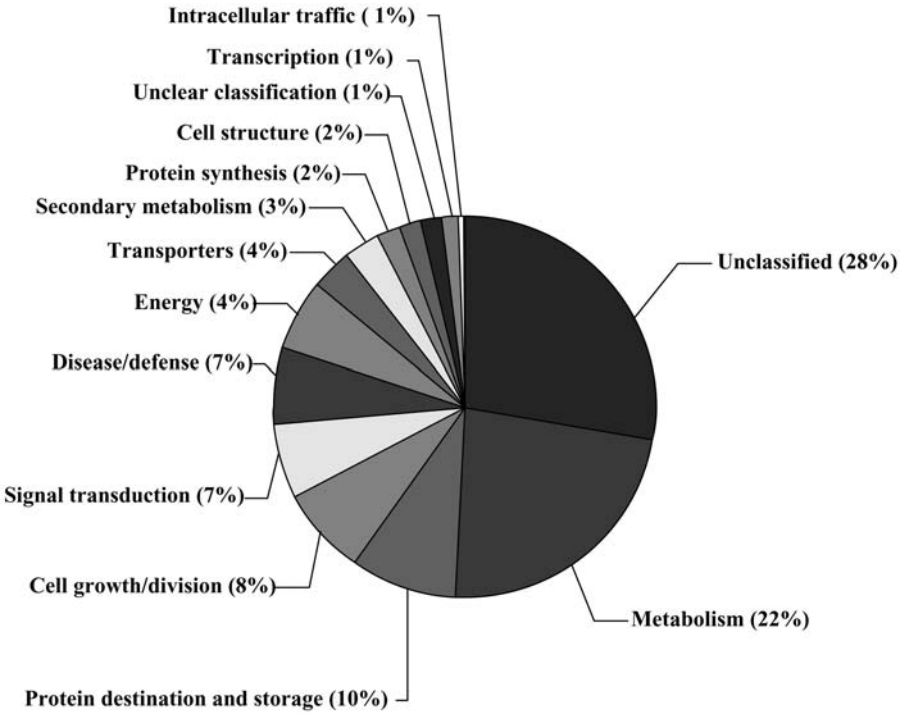

Figure 5. The functional distribution of nonredundant proteins identified from developing soybean seed. From 422 total identified proteins, 216 had nonredundant functions. The pie chart shows the distribution of these nonredundant genes after functional classification. Functional classification was based upon nomenclature by Bevan et al. (1998).

destination and storage (Fig. 6B). Eight percent of unknown proteins had stable relative abundance through the experimental period, similar to the transporter class of proteins (Fig. 6C). For the remaining 15% that had maximum abundance and 7% that had minimal abundance at 4 WAF, it is not possible to correlate function because no functional class shares these expression profiles.

Multiple Isoforms for Suc-Binding Protein and Suc Synthase Show Varied Expression Profiles

To provide carbon for all major classes of storage products, developing seeds import Suc, which is cleaved by Suc synthase (Gifford and Evans, 1981). The regulation of Suc transport throughout the plant has a tremendous impact on plant growth and productivity (Gifford and Evans, 1981). We identified 13 Suc-binding proteins mapping to the same UniGene accession number (Gma.1872). This 62-kD Suc-binding protein was previously shown to be involved in Suc transport (Grimes et al., 1992). Previous transcriptome investigation of Arabidopsis seed filling reported increased expression of Suc-binding proteins through seed filling (Ruuska et al., 2002). In this investigation, seven proteins identified as Suc-binding proteins (spots 372, 376, 383, 386, 478, 515, and 2186) increased in abundance through soybean seed filling, in agreement with the previous Arabidopsis microarray investigation. Interestingly, five Suc-binding proteins showed opposite expression trends, their relative abundance decreased during seed filling (spots 405, 463, 522, 631, and 754). The expression profiles of these proteins may support previous investigations on the importance of Suc as a signaling molecule in seed and embryo development (Konradova et al., 2002; Bate et al., 2004; Yang et al., 2004; for review, see Borisjuk et al., 2004).

In legumes, it was reported that the transcript level of Suc synthase steadily increased until the middle of development and declined thereafter. Suc levels approximately paralleled levels of Suc synthase mRNA (Heim et al., 1993). In our study, we identified six proteins (spots 225, 229, 232, 233, 234, and 237) as Suc synthases with the same UniGene number (Gma.1828). Protein expression for these proteins generally agreed with this pattern but also provided additional information about the abundance of different isoforms of Suc synthase during seed filling. For example, protein spots 225, 229, and 237 accumulated until 3 WAF, spot 234 until 4 WAF, and then all decreased. Spot 232 decreased almost linearly, and spot 233 followed this expression profile but from 5 WAF accumulated again. In contrast with Suc synthase accumulation in legumes, in Arabidopsis seeds the expression of several Suc synthase clones revealed by microarray analysis increased strongly after 4 DAF (Ruuska et al., 2002). Similar results were obtained also in developing rapeseeds (King et al., 1997). In contrast with legumes, Suc synthase accumulation in Arabidopsis and rape coincided with an increase in oil and protein synthesis rather than with the intermittent starch accumulation in young seeds (Ruuska et al., 2002).

The products of Suc synthase are largely metabolized through the glycolytic pathway, before utilization in subsequent reactions, such as amino acid and fatty acid biosynthesis. Oilseeds have a complete set

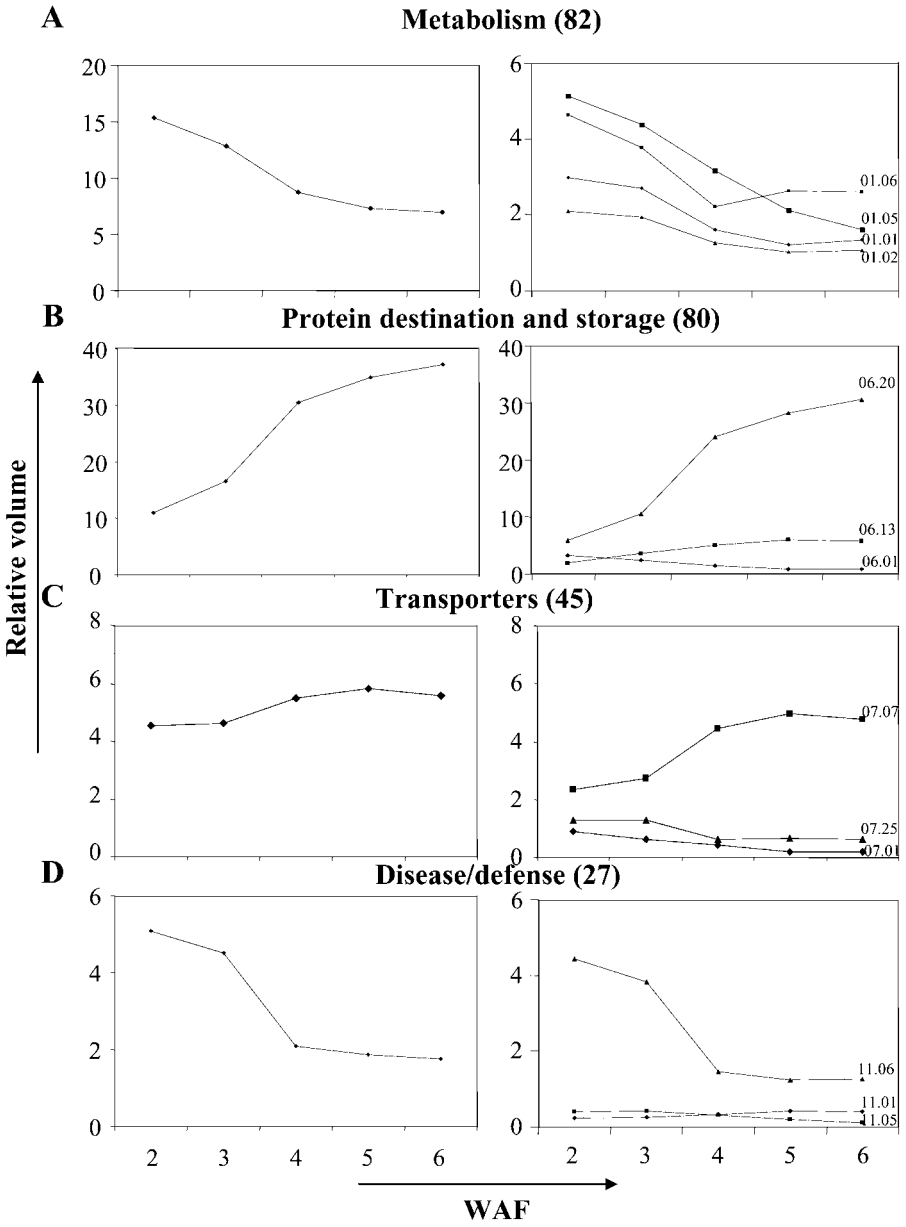

Figure 6. Composite protein expression profiles of gene function categories. The combined expression profiles were calculated as the sum of all relative volumes for each protein in the functional category. A, The combined expression profile of all 82 proteins involved in metabolism (left). Additional plot (right) shows 12 proteins involved in metabolism of amino acids (01.01), 17 in nitrogen and sulfur (01.02), 24 proteins involved in metabolism of sugars and polysaccharides (01.05), and 27 that are involved in metabolism of lipids and sterols (01.06). B, Composite expression profiles of 80 proteins involved in protein destination and storage (left). Additional plot (right) is for 12 proteins involved in folding and stability (06.01), 14 involved in proteolysis (06.13), and for 54 storage proteins (06.20). C, The combined expression profile of 45 proteins identified as transporters. More detailed view is shown by 7 ion transporters (07.01), 23 sugar transporters (07.07), and 15 ABC-type transporters (07.25). D, The combined expression profile of 27 proteins involved in disease and defense mechanisms. Additional charts show 3 resistance proteins (11.01), 4 stress responses proteins (11.05), and 20 detoxification proteins (11.06).

of glycolytic enzymes in plastids, in addition to the cytosolic pathway (Miernyk and Dennis, 1983). Interestingly, the activity of several plastic glycolytic enzymes decreased during oil synthesis in rape embryos (Eastmond and Rawsthorne, 2000). Results published on Arabidopsis seeds suggest that a major route of carbohydrate conversion into oil involves cytosolic glycolysis to phospho*enol*pyruvate (PEP), followed by PEP import to plastids and its conversion to pyruvate and acetyl-CoA for de novo fatty acid synthesis (White et al., 2000; Ruuska et al., 2002). Our results indicate the presence of a cytosolic pyruvate kinase (UniGene accession no. Gma.4073) during seed filling. This enzyme was isolated from developing soybean seeds, and the authors found 51- and 55-kD subunits (Tang et al., 2003). Our results also indicated two pyruvate kinases, spot 2270 at 54 kD and spot 2390 at 47 kD. The relative expression profiles of these spots decreased during development, which approximated the relative gene expression data for cytosolic pyruvate kinase obtained during microarray analysis of developing Arabidopsis seeds (Ruuska et al., 2002).

Key Enzymes of Cys and Met Biosynthesis Share Similar Expression Profiles

Figure 6A gives a general overview on the expression of 82 identified proteins associated with metabolism. The expression of metabolic proteins gradually decreased during seed filling. It is well known that

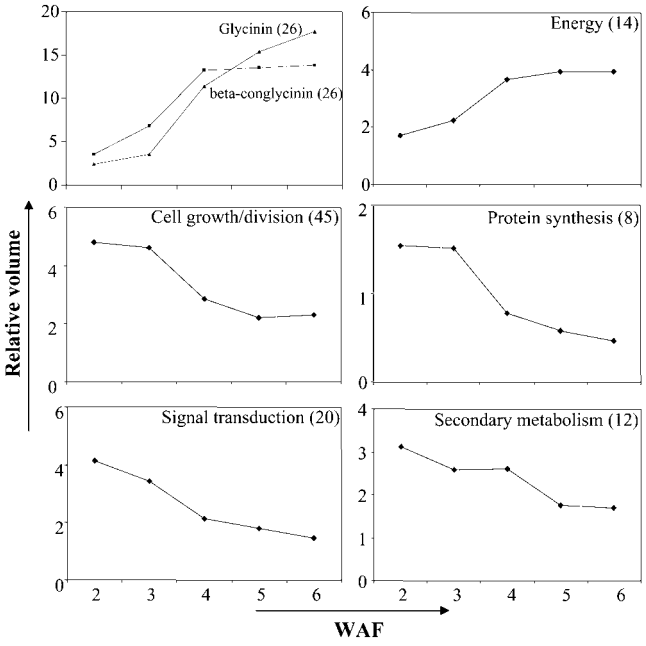

Figure 7. Composite expression profiles of various classes of proteins during soybean seed filling. The two principal soybean seed storage proteins, glycinin and β-conglycinin, both were represented by 26 spots. The figure also shows additional classes: 14 energy proteins, 45 cell growth and division proteins, 8 proteins involved in protein synthesis, 20 signal transduction proteins, and 12 secondary metabolism proteins.

soybean seeds and seed proteins are deficient in the sulfur-containing amino acids Cys and Met, and a major research effort is concentrated on increasing sulfur-containing amino acids within soybean seeds. The key enzymes for biosynthesis of both amino acids, Cys synthase (spot 984) and Met synthase (spot 281), were found in this study. The peak of expression for Cys synthase was at 3 WAF, after which the expression level decreased rapidly. In the case of Met synthase, the relative abundance rapidly decreased through seed filling and was about 5 times lower at 6 WAF when compared with 2 WAF. This is in agreement with previous studies in which Met synthase was reported to be present at low levels in mature Arabidopsis seeds (Gallardo et al., 2002).

14-3-3-Like Proteins May Perform an Important Role during Seed Development

The 14-3-3 proteins play important roles in various aspects of plant and metabolic processes (for review, see Roberts, 2003). Recent evidence suggests involvement of 14-3-3 proteins in coordination of primary carbon and nitrogen metabolism (Comparot et al., 2003). Proteomic analysis of nitrogen mobilization in pea leaves during seed filling (Schiltz et al., 2004) identified three 14-3-3-like proteins, displaying different expression patterns; two proteins increased and one decreased in abundance. Our study on soybean seed filling identified four 14-3-3-like proteins. Interestingly, three of them (spots 1027, 1029, and 1081) increased and one (spot 1046) decreased in expression during the experimental period. Since these proteins are regulatory in function, the presence of these proteins during seed filling warrants further investigation.

Lipoxygenases May Have Multiple Roles during Seed Filling Based on Contrasting Expression Profiles

Lipoxygenase (LOX) activities have been shown to be induced during plant defense responses (Keppler and Novacky, 1987). LOX also may have a direct role in signaling plant defense and wounding responses via the production of jasmonic acid (Farmer and Ryan, 1992), and many jasmonate-inducible proteins appear to be involved in plant defense responses (Reinbothe et al., 1994).

In this study, we identified 8 unique LOX representing 16 individual spots. Five spots (204, 746, 758, 982, and 2373) were identified as LOX L9 originally isolated from soybean var Paldal (Ahn et al., 2003). One of the detected LOX L9 (spot 746) was the most abundant LOX, with relative abundance of 0.7% $\pm$ 0.17% at 2 WAF. The activity of LOX L9 was reported to be down-regulated by the suppression of ascorbate peroxidase (spot 1095) and catalase (spots 418, 568, 611, 717, 812, 2297, 2332, and 2389), which are known to scavenge endogenous H_2O_2 (Ahn et al., 2003). One protein was identified as LOX vlxC, and its abundance decreased during seed filling (spot 2092). This LOX accumulates in mature leaves in response to a variety of sink limitations (Bunker et al., 1995) and was localized to the cytoplasm of a single discrete cell layer in the mesocarp of soybean pod walls (Dubbs and Grimes, 2000). Interestingly, vlxC was the second most abundant LOX, with relative abundance of 0.75% $\pm$ 0.11% at 2 WAF. Another two LOX with decreasing expression

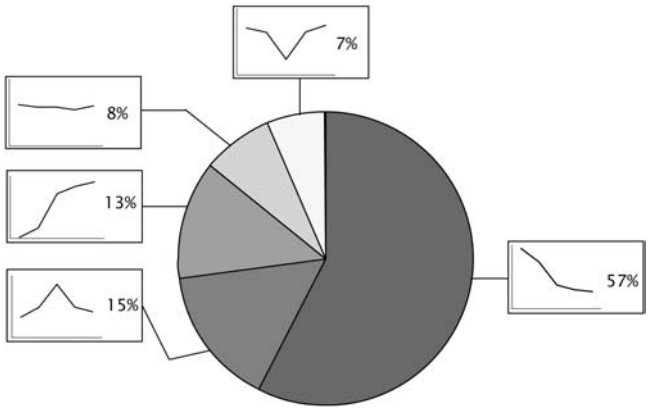

Figure 8. Expression profiles of 92 proteins annotated in the UniGene database as unknown. The expression profiles were grouped into five expression profile groups, and the percentage of unknown proteins in each group was calculated.

profiles (spots 185, 187) were isolated previously from soybean var Williams 82 (UniGene accession no. Gma.7625). Another LOX protein increased in expression during the experimental period (spot 204) and was identified as LOX 1, an enzyme involved in the metabolism of polyunsaturated fatty acids in eukaryotes (Di Venere et al., 2003). Two spots that increased during seed development were identified as LOX 2 (spots 206, 220), originally characterized by Shibata et al. (1988). LOX 3, originally identified from soybean (Kramer et al., 1994), represented by spot 212, also increased during the seed-filling period. Additionally, one protein (spot 164) with decreasing relative abundance was identified as LOX 5. Another four (spots 172, 277, 346, and 2081) with decreasing relative abundance were identified as LOX 7. One of these (spot 2081) had the third highest abundance between identified LOX, with relative abundance of 0.38% ± 0.12% at 2 WAF. This LOX has been cloned from soybean by Saravitz and Siedow (1996) and was reported to be induced by wounding of soybean leaves.

Seed Storage Proteins Are the Major Contributors to Protein Content at Later Stages of Seed Filling

The two major seed storage proteins from soybean are multimeric with Svedberg coefficients of 7 and 11S. The 7S globulins are comprised of β-conglycinin subunits, whereas 11S globulins are comprised of glycinin proteins (Roberts and Briggs, 1965; Hill and Breidenbach, 1974; Thanh and Shibasaki, 1976). This study identified 26 protein spots each as β-conglycinin and glycinin. Genetic redundancy is one possible explanation for the multiple species observed within these proteins. The β-conglycinin class of 7S seed storage proteins is a large multigene family that codes for two mRNA classes of 2.5 and 1.7 kb (Goldberg et al., 1981; Schuler et al., 1982a, 1982b; Harada et al., 1989). The 2.5-kb mRNA encodes two protein subunits, α and α′, whereas the 1.7-kb mRNA encodes the β-subunit (Beachy et al., 1981; Schuler et al., 1982b). We identified 16 protein spots under UniGene accession number Gma.2270 annotated as pre-pro β-conglycinin α′-subunits, an additional 7 protein spots under UniGene accession number Gma.17512, and 1 (Gma.5834) annotated as β-conglycinin α-subunit (Table I). Finally, two protein spots were identified (spots 638, 650) as β-conglycinin β-subunits. Regarding the second major soybean seed storage protein, our study identified 12 proteins as glycinin A2B1a precursors (Gma.1857), 2 proteins as glycinin A5A4B3 subunit precursors (Gma.1812), 7 glycinin precursors (Gma.1897), and 5 glycinins with UniGene accession number Gma.8205 (Table I). When these results were compared to previously published studies on proteomics of mature soybean seeds (Herman et al., 2003; Mooney and Thelen, 2004), nearly twice as many α-subunits of β-conglycinin were identified here, which may suggest a more complex organization than previously thought.

Soybean Allergenic Proteins Begin to Accumulate as Early as 2 WAF

Three soybean proteins, Gly m Bd 60 K, Gly m Bd 30 K, and Gly m Bd 28 K, represent the main seed allergens in soybean-sensitive patients (Ogawa et al., 2000). Gly m Bd 60 K is the α-subunit of β-conglycinin, and Gly m Bd 30 K was identified as a 34-kD seed vacuolar protein, P34 (Kalinski et al., 1990, 1992; Ogawa et al., 1993). We identified the precursor for this protein (spot 722), which started to accumulate at 2 WAF and reached its maximum abundance at 4 WAF (0.43% ± 0.038%). The position of the 34-kD seed vacuolar protein precursor (spot 722) on our gel is different when compared with the 2-DE gel reference map from a previous study of allergen proteins from soybean (Herman et al., 2003), in which the authors identified the mature 34-kD seed vacuolar protein, not the precursor. The last protein in the series, Gly m Bd 28 K, also was identified in this investigation (spot 2779). It is MP27-MP33 homolog, a minor soybean seed globulin (Tsuji et al., 1995). Accumulation of this protein began at 2 WAF, with relative abundance of 0.015% ± 0.013%, and then gradually increased to reach 0.58% ± 0.27% at 6 WAF. A better understanding of the expression profiles for the allergens may assist targeted transgenic down-regulation.

Soybean Kunitz trypsin inhibitor (SKTI) is a small monomeric, nonglycosylated protein containing 181 residues that has been characterized as a human food allergen (Burks et al., 1994; Quirce et al., 2002). Kunitz trypsin inhibitor was initially characterized by Kunitz (1945) and has been found to have six forms: *Tia* and *Tiab* (Singh et al., 1969), *Tic* (Hymowitz, 1973), *Tid* (Zhao and Wang, 1992), *Tie* (Wang et al., 1996, 2001), and null-type *ti* (Orf and Hymowitz, 1979). Our study identified four proteins (spots 1219, 1236, 1241, and 1247) as Kunitz trypsin inhibitor subtype A (UniGene accession no. Gma.13030). Each begins to accumulate at 3 WAF. According to expression profiles, it is possible to divide detected SKTI into two groups. The most abundant protein (spot 1247) accumulated almost linearly up to 5 WAF and then reached maximum at 6 WAF with a relative abundance of 2.96% ± 0.36%. The second most abundant SKTI (spot 1241) also followed this trend. The remaining two low abundant SKTI (spots 1219 and 1236) also accumulated at 3 WAF, reached a maximum at 4 WAF, and then decreased in abundance.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Soybean plants (*Glycine max* L. Maverick) were grown (ProMix soil, Quakertown, PA) in a greenhouse in Columbia, Missouri, with supplemental lighting (16-h-light/8-h-dark cycle, 26°C day/21°C night). Plants were not nodulated and were fertilized at 2-week intervals (all-purpose fertilizer, 15-30-15). Soybean flowers were tagged after opening (between 1 and 3 PM CST), and developing seed were collected precisely 2, 3, 4, 5, and 6 WAF, i.e. 14, 21, 28, 35, and 42 DAF. Seed length, thickness, and width were measured using a micrometer. Fresh and dry weight as well as total protein content were

measured at each seed stage. Protein was quantified from whole-seed homogenates using a dye-binding protein assay (Bradford, 1976).

Two-Dimensional Gel Electrophoresis

Total protein was isolated from developing seed according to a modified phenol-based procedure based upon that of Hurkman and Tanaka (1986). Immature seeds (0.5 g) were pulverized to a fine powder with liquid nitrogen and a mortar and pestle. The powder was resuspended directly in 10 mL of homogenization media (50% [v/v] phenol, 0.45 M Suc, 5 mM EDTA, 0.2% [v/v] 2-mercaptoethanol, 50 mM Tris-HCl, pH 8.8). The homogenate was allowed to reach room temperature, transferred to a 50-mL polypropylene tube, and mixed on a Nutator (Becton Dickinson, Franklin Lakes, NJ) shaken for 30 min at 4°C. After the 15-min centrifugation at 5,000*g* and 4°C, the phenol phase was removed and proteins were precipitated with 5 volumes of ice-cold 0.1 M ammonium acetate in 100% methanol at −20°C for 16 h. Then, after 10 min of centrifugation at 5,000*g*, the protein pellet was thoroughly washed twice in 20 mL of 0.1 M ammonium acetate in 100% methanol, followed by 2 washes in ice-cold 80% acetone and final wash in 70% ethanol. The protein pellet was dried for 5 min at 37°C and resuspended immediately prior to isoelectric focusing (IEF).

Proteins pellets were resuspended in 1 mL of IEF sample extraction media (8 M urea, 2 M thiourea, 2% [w/v] CHAPS, 2% [v/v] Triton X-100, 50 mM dithiothreitol) by vortexing at low speed for 1 h. Insoluble matter was removed by centrifugation for 20 min at 14,000*g*. Protein concentration was determined using a protein assay from Bio-Rad (Hercules, CA), based upon the modified procedure of Bradford (1976). Protein quantification was performed in triplicate against a standard curve of chicken γ-globulin. The desired amount of protein (1 mg) was added to a 1.5-mL tube, and volume was brought up to 450 μL with IEF extraction media. Ampholytes (2.25 μL) were added, mixed by vortexing, and centrifuged for 5 min at 14,000*g* to remove the remaining insoluble matter. The protein mix was transferred into an IEF tray, and an IPG (24 cm; Amersham Biosciences, Uppsala) was carefully placed onto the protein sample and allowed to rehydrate for 1 h prior to the addition of mineral oil. The IEF tray was placed into protean IEF cell unit (Bio-Rad). Active rehydration (10 h at 50 V) was performed, and this was directly linked to the three-step focusing protocol: 100 V for 100 Vh, 500 V for 500 Vh, and 8,000 V for 99 kVh.

Following IEF, IPG strips were removed from focusing tray, blotted to remove mineral oil, and incubated in SDS equilibration buffer (1.5 M Tris-HCl, 6 M urea, 30% [v/v] glycerol, 5% [w/v] SDS) for 15 min with 2% (w/v) dithiothreitol followed with another 15 min with 2.5% (w/v) of iodoacetamide. IPG strips were then rinsed with SDS running buffer (25 mM Tris, 0.192 M Gly, and 0.1% [w/v] SDS) and placed onto 12% acrylamide gels. Strips were then overlayed with 0.5% (w/v) agarose in SDS running buffer. Second-dimension SDS-PAGE was performed in an Ettan Dalt 12 electrophoresis unit (Amersham Biosciences, Piscataway, NJ) for 2 W/gel for 16 h or until dye migrated off the gel. Following SDS-PAGE, gels were washed in deionized water three times for 15 min and stained for a least 12 h with colloidal Coomassie (20% [v/v] ethanol, 1.6% [v/v] phosphoric acid, 8% [w/v] ammonium sulfate, 0.08% [w/v] Coomassie Brilliant Blue G-250).

Analysis of 2-D Gels

Image acquisition was performed using a ScanMaker 9800XL (Microtek, Carson, CA) with a resolution of 300 dpi and 16-bit grayscale pixel depth. Image analysis was carried out with ImageMaster 2D Platinum software version 5.0 (Amersham Biosciences, Uppsala), which allows spot detection, quantification, background subtraction, and spot matching among multiple gels. Each stage of seed development was analyzed at a minimum in biological triplicate. Protein spots were selected for profile analysis only if they were confirmed in at least three independent sample sets. For the purpose of spot analysis, a reference or pooled gel was created by pooling equal amounts of protein (200 μg) of each seed stage onto one gel.

To compensate for subtle differences in sample loading, gel staining, and destaining, the volume of each spot (i.e. spot abundance) was normalized as relative volume. This normalization method provided by ImageMaster 2D Platinum software divides each spot volume value by the sum of total spot volume values to obtain individual relative spot volumes. Total spot volume refers to the sum volume of all spots chosen for analysis; therefore, the volumes of spots excluded from the analysis were not considered. After spot detection, quantification, and background subtraction, each analyzed gel was matched individually to the reference gel, and matched spots were grouped into subclasses.

To enable direct comparison between protein relative abundances (expressed in relative volumes) from the two independent datasets (i.e. pI 4–7 and pI 3–10), a second normalization was performed. For each experimental stage, the total spot volume was calculated for each gel in the analysis separately, and the correction constants for gels pI 4 to 7 ($C_{\text{pI 4-7}}$) and pI 3 to 10 ($C_{\text{pI 3-10}}$) were calculated using following formulas:

$$C_{\text{pI 4-7}} = \frac{\sum_{S=1}^{n} \text{Vol}_S^{\text{pI 4-7}}}{\sum_{S=1}^{n} \text{Vol}_S^{\text{pI 4-7}} + \sum_{S=1}^{n} \text{Vol}_S^{\text{pI 3-10}}} \quad \text{and}$$

$$C_{\text{pI 3-10}} = \frac{\sum_{S=1}^{n} \text{Vol}_S^{\text{pI 3-10}}}{\sum_{S=1}^{n} \text{Vol}_S^{\text{pI 4-7}} + \sum_{S=1}^{n} \text{Vol}_S^{\text{pI 3-10}}},$$

where Vol_S is the volume of spot S in the gel containing n spots. The average correction constants from each experimental stage were calculated and used for correction by multiplying individual relative volumes with the correction constant for particular dataset (i.e. pI 4–7 and pI 3–10 datasets). Finally, average corrected relative volumes (from biological triplicate analyses) and SDs for each developmental stage within each protein group were calculated, and values were plotted on a line graph and deposited into a Web-based database.

Protein Identification by Mass Spectrometry

Each protein spot from a pooled reference gel was arrayed into 96-well MultiScreen plate model R5.5μm hydrophilic PTFE membrane, glass-filled polypropylene plates (Millipore, Bedford, MA) using 3.0-mm-diameter pins on the GelPix robotic spot excision station (Genetix, Milton, UK) or using a 1.5-mm-diameter manual excision pen (Gel Company, San Francisco). After arraying of protein spots, gel plugs were destained three times in 50% acetonitrile, 50 mM ammonium bicarbonate for 15 min at room temperature with gentle agitation using a microplate shaker. Destaining solution was evacuated from the bottom of the filter plates using a vacuum manifold (Millipore). Gel plugs were dehydrated in 100% acetonitrile for 5 min at room temperature. Acetonitrile was evacuated, and plates were gently blotted to filter paper to remove residual acetonitrile. Then 96-well polypropylene V-bottom sample collection plates were placed underneath each MultiScreen plate. The gel plugs were digested with 50 μL of sequencing grade trypsin manufactured by Promega (Madison, WI; 0.004 μg/μL in 50 mM ammonium bicarbonate) for 16 h at 37°C. After digestion, tryptic peptides were extracted from gel plugs using 100 μL of extraction solution (60% [v/v] acetonitrile, 1% [v/v] formic acid). After 10 min of extraction with gentle agitation, peptides were collected into V-bottom 96 collection plates by centrifugation at 2,000*g* for 2 min. The extraction procedure was repeated with the same amount of extraction solution. Extracted tryptic peptides were dried by centrifugal vacuum evaporation and stored at −80°C until analysis.

Prior to mass spectrometry, peptides were desalted using C18 microbed chromatography (ZipTip; Millipore). Dried tryptic peptides were resuspended in 10 μL of 1% formic acid. ZipTips were conditioned in 100% acetonitrile, followed by 70% (v/v) acetonitrile, 1% (v/v) formic acid and finally in 1% (v/v) formic acid. After peptide binding and washing with 1% (v/v) formic acid, peptides were eluted from C18 matrix with 7 μL of elution solution (70% [v/v] acetonitrile, 1% [v/v] formic acid). Then 0.5 μL of eluted peptides was applied to a 96 × 2 Teflon MALDI plate and mixed on target with 0.5 μL of matrix solution: 10 mg/mL α-cyano-4-hydroxycinnamic acid (Sigma-Aldrich Fluka, St. Louis) prepared in 70% (v/v) acetonitrile, 1% (v/v) formic acid and 10 mM ammonium phosphate. Analysis of trypsin-digested protein samples was carried out on Voyager-DE Pro MALDI-TOF-MS (Applied Biosystems, Foster City, CA). The MALDI-TOF was operated in the positive ion delayed extraction reflector mode for highest resolution and mass accuracy. Peptides were ionized with 337-nm laser and spectra were acquired at 20 kV acceleration potential with optimized parameters. Close external calibration was employed using a mixture of standard peptides (Applied Biosystems). This calibration method typically provided mass accuracy of 25 to 75 ppm across the mass range 700 to 4,500 D. Peptide spectra were automatically processed for baseline correction, noise removal, peak deisotoping, and threshold adjustment (2% base peak intensity) prior to submission to a local copy of version 3.2.1 of the MS-Fit program of Protein Prospector (http://prospector.ucsf.edu) through the Proteomics Solution I software suite (Applied Biosystems). Search criteria required the match of at least four peptides with a mass error of less than 150 ppm for tentative protein assignment. All searches were performed against the National Center for Biotechnology Information (NCBI) soybean UniGene database

(12,629 entries as of August 2004; http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene). Candidate identifications with MOWSE scores higher than 1×10^6 were automatically considered as positive assignments. For all other assignments of protein spots, the MOWSE score cutoff threshold was 1×10^3. Additionally, positive protein assignments required greater than 4% sequence coverage and less than 25% deviation between theoretical and experimental MW and pI values obtained from calibrated 2-D gels. If more than one protein satisfied mentioned threshold criteria, the entry with the highest MOWSE score was assigned.

Web Database Construction

Programming for the Web database was scripted through C (Kerninghan and Richie, 1978) and gnuplot (www.gnuplot.info). All expression data were parsed and formatted prior to creating expression graphs using the gnuplot software package. Simple script in C automated the process of creating expression pages for each spot. Scripts were used to create mass spectrometry data files from the analysis data. Spot coordinates along with spot identification were exported from ImageMaster software to XML Web page. This XML page was parsed, and coordinates for each spot identification were transformed according to the image resolution and tagged to the expression page through standard scripting. Gcc-3.4 compiler was used to compile the scripts, which were executed on a Linux 2.4.20-based server.

Supplemental Data

A proteomics database archiving all protein expression and identification data discussed here can be found at http://oilseedproteomics.missouri.edu/. The supplemental files online contain all expression profile data.

ACKNOWLEDGMENTS

We thank Alan Marshall for database maintenance and server space. We also thank Professor Dong Xu (University of Missouri, Columbia) for providing computational resources for script writing and data processing.

Received November 16, 2004; returned for revision December 29, 2004; accepted December 30, 2004.

LITERATURE CITED

Ahn BM, Kim YK, Kim SY, Kim MW, Oh SE (2003) Suppression of lipoxygenase by hydrogen peroxide in soybean (*Glycine max* L cv PALDAL) suspension cells. Korean J Genet **25:** 251–259

Bate NJ, Niu XP, Wang YW, Reimann KS, Helentjaris TG (2004) An invertase inhibitor from maize localizes to the embryo surrounding region during early kernel development. Plant Physiol **134:** 246–254

Beachy RN, Jarvis NP, Barton KA (1981) Biosynthesis of subunits of the soybean 7S storage protein. J Mol Appl Genet **1:** 19–27

Bevan M, Bancroft I, Bent E, Love K, Goodman H, Dean C, Bergkamp R, Dirkse W, Van Staveren M, Stiekema W, et al (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana.* Nature **391:** 485–488

Borisjuk L, Rolletschek H, Radchuk R, Weschke W, Wobus U, Weber H (2004) Seed development and differentiation: a role for metabolic regulation. Plant Biol **6:** 375–386

Bradford M (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Chem **72:** 248–254

Bunker TW, Koetje DS, Stephenson LC, Creelman RA, Mullet JE, Grimes HD (1995) Sink limitation induces the expression of multiple soybean vegetative lipoxygenase mRNAs while the endogenous jasmonic acid level remains low. Plant Cell **7:** 1319–1331

Burks AW, Cockrell G, Connaughton C, Helm RM (1994) Epitope specificity and immunoaffinity purification of the major peanut allergen, Ara h I. J Allergy Clin Immunol **93:** 743–750

Comparot S, Lingiah G, Martin T (2003) Function and specificity of 14-3-3 proteins in the regulation of carbohydrate and nitrogen metabolism. J Exp Bot **54:** 595–604

Di Venere A, Salucci ML, van Zadelhoff G, Veldink G, Mei G, Rosato N, Finazzi-Agro A, Maccarrone M (2003) Structure-to-function relationship of mini-lipoxygenase, a 60-kDa fragment of soybean lipoxygenase-1 with lower stability but higher enzymatic activity. J Biol Chem **278:** 18281–18288

Dubbs WE, Grimes HD (2000) Specific lipoxygenase isoforms accumulate in distinct regions of soybean pod walls and mark a unique cell layer. Plant Physiol **123:** 1269–1279

Eastmond PJ, Rawsthorne S (2000) Coordinate changes in carbon partitioning and plastidial metabolism during the development of oilseed rape embryo. Plant Physiol **122:** 767–774

Farmer EE, Ryan CA (1992) Octadecanoid precursors of jasmonic acid activate the synthesis of wound-inducible proteinase inhibitors. Plant Cell **4:** 129–134

Gallardo K, Job C, Groot SP, Puype M, Demol H, Vandekerckhove J, Job D (2002) Proteomics of Arabidopsis seed germination. A comparative study of wild-type and gibberellin-deficient seeds. Plant Physiol **129:** 823–837

Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD, Burstin J (2003) Proteomics of *Medicago truncatula* seed development established the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol **133:** 664–682

Gifford RM, Evans RT (1981) Photosynthesis, carbon partitioning, and yield. Annu Rev Plant Physiol **32:** 485–509

Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J (2000) Microarray analysis of developing *Arabidopsis* seeds. Plant Physiol **124:** 1570–1581

Grimes HD, Overvoorde PJ, Ripp K, Franceschi VR, Hitz WD (1992) A 62-kD sucrose binding protein is expressed and localized in tissues actively engaged in sucrose transport. Plant Cell **4:** 1561–1574

Goldberg RB, Hoschek RB, Ditta G, Breidenbach RW (1981) Developmental regulation of cloned superabundant embryo mRNAs in soybean. Dev Biol **83:** 218–231

Gunstone FD (2001) Soybeans pace boost in oilseed production. Inform **11:** 1287–1289

Harada JJ, Barker SJ, Goldberg RB (1989) Soybean beta-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. Plant Cell **1:** 415–425

Heim U, Weber H, Baumlein H, Wobus U (1993) A sucrose synthase gene of *Vicia faba* L.: expression pattern in developing seeds in relation to starch synthesis and metabolic regulation. Planta **191:** 394–401

Herman EM, Larkins BA (1999) Protein storage bodies and vacuoles. Plant Cell **11:** 601–613

Herman EM, Helm RM, Jung R, Kinney AJ (2003) Genetic modification removes an immunodominant allergen from soybean. Plant Physiol **132:** 36–43

Hill JE, Breidenbach RW (1974) Proteins of soybean seeds. II. Accumulation of the major protein components during seed development and maturation. Plant Physiol **53:** 747–751

Hurkman WJ, Tanaka CK (1986) Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. Plant Physiol **81:** 802–806

Hymowitz T (1973) Electrophoretic analysis of SBTI-A2 in the USDA soybean germplasm collection. Crop Sci **13:** 420–421

Kalinski AJ, Melroy DL, Dwivedi RS, Herman EM (1992) A soybean vacuolar protein (P34) related to thiol proteases which is synthesized as a glycoprotein precursor during seed maturation. J Biol Chem **267:** 12068–12076

Kalinski AJ, Weisemann J, Matthews BF, Herman EM (1990) Molecular cloning of a protein associated with soybean oil bodies which is homologous to thiol proteases of the papain family. J Biol Chem **265:** 13843–13848

Keppler LD, Novacky A (1987) The initiation of membrane lipid peroxidation during bacteria-induced hypersensitive reactions. Physiol Mol Plant Pathol **30:** 233–245

Kerninghan BW, Richie DM (1978) The C programming language. Prentice-Hall Software Press, Englewood Cliffs, NJ

King SP, Lunn JE, Furbank RT (1997) Carbohydrate content and enzyme metabolism in developing canola siliques. Plant Physiol **114:** 153–160

Konradova H, Lipavska H, Albrechtova J, Vreugdenhil D (2002) Sucrose metabolism during somatic and zygotic embryogeneses in Norway spruce: content of soluble saccharides and localisation of key enzyme activities. J Plant Physiol **159:** 387–396

Kramer JA, Johnson KR, Dunham WR, Sands RH, Funk MO Jr (1994)

Position 713 is critical for catalysis but not iron binding in soybean lipoxygenase 3. Biochemistry **33:** 15017–15022

Kunitz M (1945) Crystallization of trypsin inhibitor from soybean. Science **101:** 668–669

Mienke DW, Chen J, Beachy RN (1981) Expression of storage-protein genes during soybean seed development. Planta **153:** 130–139

Miernyk JA, Dennis DT (1983) Mitochondrial, plastid, and cytosolic isozymes of hexokinase from developing endosperm of Ricinus communis. Arch Biochem Biophys **226:** 458–468

Mooney BP, Thelen JJ (2004) High-throughput peptide mass fingerprinting of soybean seed protein: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. Phytochemistry **65:** 1733–1744

Ogawa T, Samoto M, Takahashi K (2000) Soybean allergens and hypoallergenic soybean products. J Nutr Sci Vitaminol **46:** 271–279

Ogawa T, Tsuji H, Kitamura K, Zhu UL, Hirano H, Nishikawa K (1993) Identification of the soybean allergenic protein, Gly m Bd 30 K, with the soybean seed 34-kDa oil-body-associated protein. Biosci Biotechnol Biochem **57:** 1030–1033

Orf JH, Hymowitz T (1979) Inheritance of the absence of the Kunitz trypsin inhibitor in seed protein of soybeans. Crop Sci **19:** 107–109

Quirce S, Fernandez-Nieto M, Polo F, Sastre J (2002) Soybean trypsin inhibitor is an occupational inhalant allergen. J Allergy Clin Immunol **109:** 178

Pappin DJC, Hojrup P, Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol **3:** 327–332

Reinbothe S, Reinbothe C, Lehmann J, Becker W, Apel K, Parthier B (1994) JIP60, a methyl jasmonate-induced ribosome-inactivating protein involved in plant stress reactions. Proc Natl Acad Sci USA **91:** 7012–7016

Roberts MR (2003) 14-3-3 proteins find new partners in plant cell signaling. Trends Plant Sci **8:** 218–223

Roberts RC, Briggs DR (1965) Isolation and characterization of the 7S component of soybean globulins. Cereal Chem **42:** 71–85

Rubel A, Rinne W, Canvin DT (1972) Protein oil and fatty acid in developing soybean seeds. Crop Sci **12:** 739–741

Ruuska SS, Girke T, Benning C, Ohlrogge JB (2002) Contrapuntal networks of gene expression during *Arabidopsis* seed filling. Plant Cell **14:** 1191–1206

Saravitz DM, Siedow JN (1996) The differential expression of wound-inducible lipoxygenase genes in soybean leaves. Plant Physiol **110:** 287–299

Schiltz S, Gallardo K, Huart M, Negrono L, Sommerer N, Burstin J (2004) Proteome reference maps of vegetable tissue in pea. An investigation of nitrogen mobilization from leaves during seed filling. Plant Physiol **135:** 2241–2260

Schuler MA, Ladin BF, Fryer G, Pollaco JG, Beachy RN (1982b) Structural sequences are conserved in the genes coding for the α, α' and β subunits of the soybean seed storage protein and other seed proteins. Nucleic Acids Res **10:** 8245–8260

Schuler MA, Schmitt ES, Beachy RN (1982a) Closely related families of genes code for the alpha and alpha′ subunits of the soybean 7S storage protein complex. Nucleic Acids Res **10:** 8225–8244

Shibata D, Steczko J, Dixon JE, Andrews PC, Hermodson M, Axelrod B (1988) Primary structure of soybean lipoxygenase L-2. J Biol Chem **263:** 6816–6821

Singh L, Wilson CM, Hadley HH (1969) Genetic differences in soybean trypsin inhibitor separated by disk electrophoresis. Crop Sci **9:** 489–491

Tang GQ, Hardin SC, Dewey R, Huber SC (2003) A novel C-terminal proteolytic processing of cytosolic pyruvate kinase, its phosphorylation and degradation by the proteasome in developing soybean seeds. Plant J **34:** 77–93

Thanh VH, Shibasaki K (1976) Heterogeneity of beta-conglycinin. Biochim Biophys Acta **181:** 404–409

Thelen JJ, Ohlrogge JB (2002) Metabolic engineering of fatty acid biosynthesis in plants. Metab Eng **4:** 12–21

Tsuji H, Okada N, Yamanishi R, Bando N, Kimoto M, Ogawa T (1995) Measurement of Gly m Bd 30 K, a major soybean allergen, in soybean products by a sandwich enzyme-linked immunosorbent assay. Biosci Biotechnol Biochem **59:** 150–151

Wang KJ, Kaizuma N, Takahata Y, Hatakeyama S (1996) Detection of two new variants of soybean Kunitz trypsin inhibitor through electrophoresis. Breed Sci **46:** 39–44

Wang KJ, Takahata Y, Ito K, Zhao YP, Tsutsumi K, Kaizuma N (2001) Genetic characterization of a novel soybean Kunitz trypsin inhibitor. Breed Sci **51:** 185–190

White JA, Todd J, Newman T, Focks N, Girke T, Martínez de Ilárduya O, Jaworski JG, Ohlrogge JB, Benning C (2000) A new set of *Arabidopsis* expressed sequence tags from developing seeds: the metabolic pathway from carbohydrates to seed oil. Plant Physiol **124:** 1582–1594

Yang Z, Zhang L, Diao F, Huang M, Wu N (2004) Sucrose regulates elongation of carrot somatic embryo radicles as a signal molecule. Plant Mol Biol **54:** 441–459

Yazdi-Samadi B, Rinne RW, Seif RD (1977) Components of developing soybean seeds: oil protein sugars starch organic acids and amino acids. Agron J **69:** 481–486

Zhao SW, Wang H (1992) A new electrophoretic variant of SBTi-A2 in soybean seed protein. Soybean Genetics Newsletter **19:** 22–24