

# A Two-dimensional Electrophoresis Proteomic Reference Map and Systematic Identification of 1367 Proteins from a Cell Suspension Culture of the Model Legume *Medicago truncatula*\*<sup>§</sup>

Zhentian Lei‡, Aaron M. Elmer‡, Bonnie S. Watson‡, Richard A. Dixon‡, Pedro J. Mendes§, and Lloyd W. Sumner‡¶

The proteome of a *Medicago truncatula* cell suspension culture was analyzed using two-dimensional electrophoresis and nanoscale HPLC coupled to a tandem Q-TOF mass spectrometer (QSTAR Pulsar i) to yield an extensive protein reference map. Coomassie Brilliant Blue R-250 was used to visualize more than 1661 proteins, which were excised, subjected to in-gel trypsin digestion, and analyzed using nanoscale HPLC/MS/MS. The resulting spectral data were queried against a custom legume protein database using the MASCOT search engine. A total of 1367 of the 1661 proteins were identified with high rigor, yielding an identification success rate of 83% and 907 unique protein accession numbers. Functional annotation of the *M. truncatula* suspension cell proteins revealed a complete tricarboxylic acid cycle, a nearly complete glycolytic pathway, a significant portion of the ubiquitin pathway with the associated proteolytic and regulatory complexes, and many enzymes involved in secondary metabolism such as flavonoid/isoflavonoid, chalcone, and lignin biosynthesis. Proteins were also identified from most other functional classes including primary metabolism, energy production, disease/defense, protein destination/storage, protein synthesis, transcription, cell growth/division, and signal transduction. This work represents the most extensive proteomic description of *M. truncatula* suspension cells to date and provides a reference map for future comparative proteomic and functional genomic studies of the response of these cells to biotic and abiotic stress. *Molecular & Cellular Proteomics* 4:1812–1825, 2005.

Legumes (Fabaceae) are one of the most economically important crop families in the world and are characterized by

their unique ability to fix atmospheric nitrogen through a symbiotic relationship with soil bacteria. These bacteria, collectively termed Rhizobia and which include genera such as *Rhizobium*, *Sinorhizobium*, *Bradyrhizobium*, *Mesorhizobium*, and *Azorhizobium*, form specialized organs within the plant, and nitrogen fixation occurs via the conversion of N<sub>2</sub> into NH<sub>3</sub> by bacterial nitrogenases. The mutualistic interactions provide a plentiful supply of nitrogen to the plants that in turn results in very high protein levels in legumes. Therefore, legumes have been assimilated as a major dietary source of protein for both humans and animals. Legumes also provide nitrogen to the soil, thus reducing the need for exogenous fertilizers. Legumes are also a unique source of natural products such as isoflavonoids, alkaloids, and saponins, many of which have documented antimicrobial and pharmacological properties (1–3). Based on the above traits, legumes have significant economic and ecological value. For example, soybeans provide more than one-third of human protein intake and approximately half of the world's supply of oilseed. Over 73 million acres of soybeans (*Glycine max*) and 76 million acres of alfalfa (*Medicago sativa*) were cultivated in the United States in 2003 and have estimated values of \$17.5 billion and \$7.5 billion, respectively (4).

The mutualistic interactions between legumes and microbes represent an important aspect of plant biology, which cannot be studied using the model plant *Arabidopsis thaliana* as it is unable to establish root endosymbiosis with rhizobia. The use of agriculturally important legume crops such as soybean and alfalfa to dissect legume biology is complicated by their large complex polyploid genomes, highly repetitive DNA, and outcrossing nature. However, *Medicago truncatula* (common name barrel medic and a close relative to alfalfa) has emerged over the past decade as a model system for molecular and genetic studies of legume biology (5). This is mainly because of its relatively small diploid genome, autogamous pollination, and the ability to regenerate via somatic embryogenesis (6–9). Large numbers of expressed sequence

From the ‡Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, Oklahoma 73402 and the §Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Received, April 8, 2005, and in revised form, July 22, 2005

Published, MCP Papers in Press, July 26, 2005, DOI 10.1074/mcp.D500005-MCP200

This is an open access article under the CC BY license.

tag (EST)<sup>1</sup> libraries have been sequenced for *M. truncatula* and for other legumes such as *Lotus japonicus* and *G. max* (soybean). As of March 2005, 226,923 EST sequences (36,878 unique) for *M. truncatula*, 109,618 EST sequences (28,460 unique) for *L. japonicus*, and 330,436 EST sequences (63,676 unique) for *G. max* have been generated ([www.tigr.org/tdb/tgi/plant.shtml](http://www.tigr.org/tdb/tgi/plant.shtml)). Furthermore a physical map of the *M. truncatula* genome is now available (10, 11), and sequencing of the gene-rich regions of the *M. truncatula* genome is ongoing at the University of Oklahoma and laboratories in the United Kingdom and France ([medicago.org/genome/](http://medicago.org/genome/)) (12). Currently over 1000 bacterial artificial chromosomes have been sequenced yielding ~130 Mb of genomic sequence. The continuously emerging nuclear genomic sequence is complimented by an already complete 124-kb chloroplast genome sequence ([www.ncbi.nlm.nih.gov/genomes/framiki.cgi?db=genome&gi=15787](http://www.ncbi.nlm.nih.gov/genomes/framiki.cgi?db=genome&gi=15787)).

The above genomic resources constitute a critical infrastructure that is propelling *M. truncatula* forward as a model for the study of legume biology. Unfortunately this information alone is insufficient to address complex biological questions; thus, functional genomics and systems biology approaches have evolved to characterize gene function and system responses at the transcript, protein, and metabolite levels (13, 14). High density microarray analyses provide quantitative and qualitative data for large numbers of mRNAs (15, 16); however, these do not always directly reflect active enzyme levels (17). Enzyme activity is often regulated through post-translational modifications such as phosphorylation and acylation or through protein sorting, protein-protein interactions, and/or controlled proteolysis (18). Therefore the direct evaluation of proteins and the application of integrated systems approaches are highly advantageous for understanding the functional consequences of gene expression as documented in several recent journal issues focused on these topics (19–21).

Currently one of the most utilized approaches in proteomics includes 2-DE and mass spectrometry to quantify and identify proteins (22). Recent developments, including improved protein solubilization, enhanced 2-DE resolution, and increased capabilities of modern mass spectrometry to sequence peptides, have made proteomics a critical research tool in biological sciences (for reviews, see Refs. 23–26). The proteomes of yeast (27, 28), humans (29, 30), animals (31, 32), bacteria (33, 34), and plants (for reviews, see Refs. 35 and 36) have been studied to better understand their respective physiology.

Successful applications of proteomic approaches have

been previously reported in *M. truncatula* (37–39). Bestel-Corre *et al.* (37) used 2-DE and MALDI-TOF-MS to visualize and identify ~500 proteins associated with the symbiotic relationships between *M. truncatula* roots and the mycorrhizal fungus *Glomus mosseae* or the nitrogen-fixing bacterium *Sinorhizobium meliloti*. Temporal changes in the *M. truncatula* seed proteome during development were described by Gallardo *et al.* (38). A study of organ/tissue-specific proteomes from *M. truncatula* successfully identified 304 proteins in a survey of roots, leaves, stems, flowers, seed pods, and suspension cultures (39). Furthermore Imin *et al.* (40) have studied the proteome of embryogenic tissue cultures.

The objective of this work was to generate an extensive protein reference map for *M. truncatula* cell suspension cultures that will be utilized later in an ongoing integrated functional genomic study of biotic and abiotic stress responses. This report presents a rigorously characterized proteomic reference map of *M. truncatula* cell suspension cultures and represents the most comprehensive study of the *M. truncatula* proteome to date.

#### EXPERIMENTAL PROCEDURES

**Chemicals**—HPLC grade ACN and water were purchased from Burdick & Jackson (Muskegon, MI). Formic acid, iodoacetamide, and DTT were purchased from Sigma. CHAPS was purchased from United States Biochemical Corp. (Cleveland, OH).

**Tissue Culture, Protein Extraction, and 2-DE Separation**—Cell suspension cultures were initiated from *M. truncatula* cv. Jemalong A17 root calli and grown in the dark in modified Schenk and Hildebrandt (SH) liquid medium with transfer to fresh medium every 2 weeks (41, 42). Cells were harvested during the log phase of the sixth passage, washed with fresh SH medium and SH:water (1:1, v/v), and stored at –80 °C. Frozen cells were ground in liquid nitrogen and extracted with buffer containing 40 mM Tris-HCl (pH 9.5), 50 mM MgCl<sub>2</sub>, 2% (w/v) polyvinylpyrrolidone (Sigma), 1 mM phenylmethylsulfonyl fluoride, and 120 units/ml endonuclease (catalog number E8263, Sigma) followed by sonication as described previously (43). The extracts were centrifuged at 5000 × *g* for 10 min at 4 °C, and the supernatant was recovered. The supernatant was brought to a final concentration of 12.5% (w/v) TCA plus 1% β-mercaptoethanol and incubated at –20 °C for 45 min. Protein was recovered by centrifugation at 15,000 × *g* for 20 min. The protein pellets were washed with a cold solution of 80% acetone and 20% water containing 0.05% (v/v) 2-mercaptoethanol three times to remove residual TCA, air-dried, and resuspended in 2-DE solubilization buffer consisting of 9 M urea, 3% CHAPS, 2% Triton X-100, 20 mM DTT, and 0.5% ampholytes. The protein concentration was then quantified by the Bradford method (44) using a commercial dye reagent (Bio-Rad) and bovine serum albumin as the standard.

IPG strips (Immobiline™ Dry Strips, 24 cm, pH 3–10 non-linear, Amersham Biosciences) were passively rehydrated overnight with protein solution (1900 μg of protein in 450 μl). IEF of proteins was performed using the following step gradient: 500 volts for 1 h, 1000 volts for 1 h, and 8000 volts until a total of 67,500 V-h had been achieved. After IEF, strips were equilibrated in buffer containing 7 M urea, 2% SDS, 375 mM Tris (pH 8.8), and 10% glycerol plus either 50 mM DTT for reduction or 100 mM iodoacetamide for alkylation. Equilibrated IPG strips were loaded onto a 10% T acrylamide gel, sealed with 1% agarose, and electrophoresed overnight at 2 watts/gel. Gels were stained with Coomassie Brilliant Blue R-250, and images were

<sup>1</sup> The abbreviations used are: EST, expressed sequence tag; 2-DE, two-dimensional electrophoresis; MOWSE, molecular weight search; SH, Schenk and Hildebrandt; IDA, information-dependent acquisition; CHI, chalcone-flavonone isomerase; MMP, matrix metalloproteinase; E1, ubiquitin-activating enzyme; E2, ubiquitin-conjugating enzyme; E3, ubiquitin-protein ligase; E4, ubiquitination factor; PR, pathogenesis-related.

acquired on a UMax Astra 2400S scanner at 300 dpi and saved as a gray scale TIFF file. Experimental molecular weight and pI values were calculated from digitized images using molecular weight marker proteins and the predicted non-linear pH gradient provided by Amersham Biosciences. Protein spots were detected and numbered with Genomic Solutions HT Analyzer software (Genomic Solutions, Ann Arbor, MI).

**In-gel Trypsin Digestion**—A total of 1661 Coomassie Brilliant Blue-visualized protein spots, including seven positive controls (molecular weight markers), were manually excised (The Gel Company, San Francisco, CA) as 1.5- or 3.0-mm diameter plugs depending on the relative abundance of the spot. Gel plugs were transferred to polypropylene 96-well plates, sealed, and stored at  $-80^{\circ}\text{C}$  until further processing. To each well, 25  $\mu\text{L}$  of a 1:1 (v/v) solution of 50 mM ammonium bicarbonate and ACN was added, and the mixtures were incubated at room temperature for 15 min. This process was repeated until all gel spots were completely destained. The spots were then dehydrated with 25  $\mu\text{L}$  of ACN for 15 min at room temperature. After ACN removal, the gel spots were dried under vacuum and rehydrated in 20  $\mu\text{L}$  of sequencing grade modified bovine trypsin (10 ng/ $\mu\text{L}$  in 25 mM ammonium bicarbonate, Roche Diagnostics). After rehydration for 20 min, excess trypsin solution was removed, and 15  $\mu\text{L}$  of 25 mM ammonium bicarbonate was added to each well to prevent dehydration during incubation. Proteolysis was allowed to continue overnight at  $37^{\circ}\text{C}$  and stopped by adding 15  $\mu\text{L}$  of 10% formic acid. The supernatant was recovered, and the spots were extracted twice more with 25  $\mu\text{L}$  of a 1:1 (v/v) solution of ACN and 25 mM ammonium bicarbonate and once more with 25  $\mu\text{L}$  of ACN. The extracts were then combined and concentrated under vacuum to a final volume of 25  $\mu\text{L}$ .

**LC/MS/MS**—Separations of the protein digests were achieved using a nanoscale HPLC system (LC Packings, San Francisco, CA) consisting of an autosampler (Famos), a precolumn switching device (Switchos), and an HPLC pump system (Ultimate). Samples (5  $\mu\text{L}$ ) were loaded onto a  $\text{C}_{18}$  precolumn (0.3-mm inner diameter  $\times$  1.0 mm, 100  $\text{\AA}$ , PepMap  $\text{C}_{18}$ , LC Packings) for desalting and concentrating at a flow rate of 50  $\mu\text{L}/\text{min}$  using mobile phase A (5% ACN and 95% water containing 0.1% formic acid). Peptides were then eluted from the precolumn and separated on a nanoanalytical  $\text{C}_{18}$  column (75- $\mu\text{m}$  inner diameter  $\times$  15 cm, 100  $\text{\AA}$ , PepMap  $\text{C}_{18}$ , LC Packings) at a flow rate of 200 nL/min. Peptides were eluted with a linear gradient of 5–40% mobile phase B (95% ACN and 5% water containing 0.08% formic acid) over 40 min.

The separated peptides were directly analyzed with an ABI QSTAR Pulsar *i* hybrid Q-TOF mass spectrometer (Applied Biosystems) equipped with a nanoelectrospray ionization source (Protana). The nanoelectrospray was generated using a PicoTip needle (10- $\mu\text{m}$  inner diameter, New Objectives, Woburn, MA) maintained at a voltage of 2400 V. TOF-MS and tandem mass spectral data were acquired using information-dependent acquisition (IDA), charge state selection from 2 to 5, an intensity threshold of 10 counts/s, and a collision energy setting automatically determined by the IDA based on the  $m/z$  values of each precursor ion. Following IDA data acquisition, precursor ions were excluded for 90 s using a window of 6 amu to minimize the redundancy in tandem mass spectra.

**Database Queries and Protein Identifications**—The acquired mass spectral data were queried against a custom legume protein database using the MASCOT (version 1.8.0, Matrix Science Ltd., London, UK) search engine (45, 46), a mass tolerance of 150 ppm, and allowance for up to one trypsin miscleavage and variable amino acid modifications consisting of methionine oxidation and cysteine carbamidomethylation. The custom protein legume database was generated from tentative consensus sequences compiled as gene indices by The Institute for Genomic Research ([www.tigr.org/tdb/tgi.shtml](http://www.tigr.org/tdb/tgi.shtml)). These sequences included MtGl.053002 (*M. truncatula*, 22,652 records),

GmGl.052802 (*G. max*, 32,081 records), and LjGl.053102 (*L. japonica*, 7,686 records). The nucleotide sequences were then translated into amino acid sequences (62,651 records) and annotated using EST Analyzer ([bioinfo.noble.org/](http://bioinfo.noble.org/)). For a given sequence, EST Analyzer searched the National Center for Biotechnology nonredundant (NCBI nr) protein dataset to identify a homologous protein (the best hit of BLASTX search), which was used to annotate the query sequence. Based on the alignment between the query sequence and template, frameshift errors were also detected and corrected if possible. All possible protein sequences were annotated, given pseudo-GI numbers, formatted similarly to NCBI nr to allow queries by MASCOT, and compiled as a plant protein database. The dataset also included *M. truncatula* chloroplast sequences (MtChl v.1, 156 records) and *Arabidopsis* mitochondrial proteins (AtMit v.1, 46 records). Only protein identifications with a molecular weight search (MOWSE) score (45, 46) greater than 2 times the generally accepted significant threshold (determined at 95% confidence level as calculated by MASCOT) and at least two peptides matched are reported in this study.

## RESULTS AND DISCUSSION

**2-DE Protein Reference Map and Protein Identification**—The proteome of *M. truncatula* cell suspension culture initiated from roots was profiled using 2-DE. Multiple 2-DE gels were acquired, and the best gel was selected to serve as the reference map (Fig. 1). Analytical and biological variances were calculated for triplicate 2-DE gels obtained for similar but independent culture flasks. The average coefficient of variance (*i.e.* relative standard deviation) for the total number of protein spots detected was determined to be 6%, and the variance in normalized spot volume was determined to be 48%. These values are similar to those reported by others in the literature for 2-DE (47, 48).

The reference map contained 1661 protein spots and was populated primarily by acidic ( $\text{pI} < 7$ ) proteins but also contained a limited number of basic proteins. The vast majority of proteins (~80%) visualized possessed molecular masses between 22 and 97 kDa. Only 7.7% of the proteins had molecular masses greater than 97 kDa, and 12.3% had molecular masses lower than 22 kDa. This is in agreement with Candiano *et al.* (49) who reported that current commercially available 4% total acrylamide IPG gel strips limit the high molecular weight proteins visualized with 2-DE.

All 1661 protein spots were excised, digested in-gel with trypsin, and analyzed by nano-HPLC coupled to an ABI hybrid quadrupole time-of-flight mass spectrometer (ABI QSTAR Pulsar *i*). Representative LC/MS/MS data, an experimentally determined peptide sequence, and a database search result are shown in Fig. 2. To significantly elevate the confidence in the reported protein identifications, a score criterion of twice the normal MASCOT score considered significant at  $p \leq 0.05$  was used (*i.e.* score  $> \sim 80$ ). Using this criterion, proteins from 1367 of the 1661 spots were identified, yielding a protein identification success rate of 83%. This is a substantial increase as compared with the previously reported protein identification success rates for *M. truncatula* of 37% (50) and 55% (39) using peptide mass fingerprinting. The improvement is attributed to greater protein coverage obtained by LC/



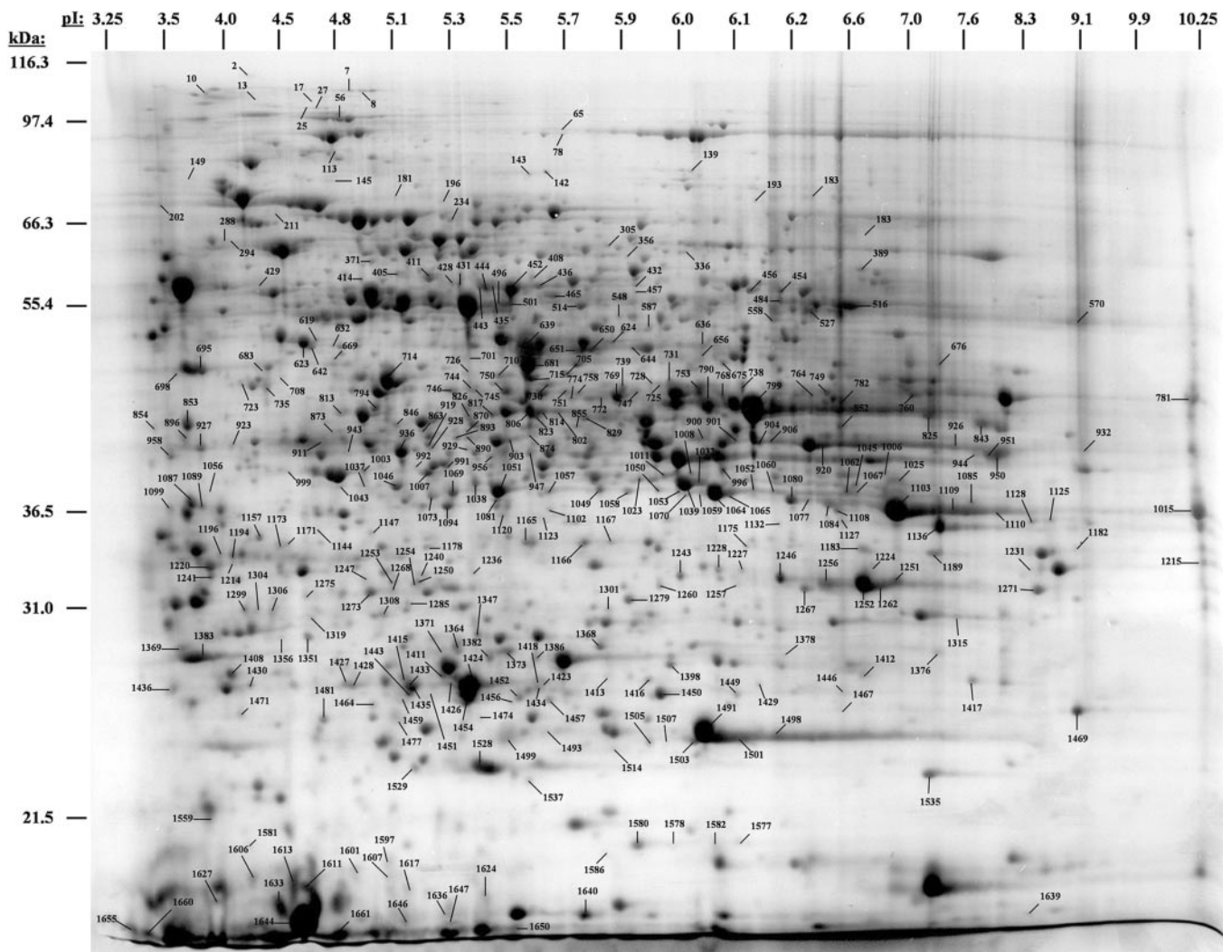


FIG. 1. 2-DE proteome reference maps of *M. truncatula* suspension cells cultures. 2-DE was performed using 1900  $\mu$ g of protein, non-linear 24-cm IPG strips (pH 3–10), and a 10% (w/v) total acrylamide SDS second dimension. Proteins involved in secondary metabolism and disease/defense are numbered.

MS/MS and the use of the custom legume database. A comprehensive list of all identified proteins, their accession numbers, number of peptides matched, sequence coverage, MOWSE scores, experimental pI values, and molecular weights are provided in the supplemental materials (Supplemental Table 1). Additionally an annotated reference gel image is available upon request. These data establish the high level of analytical rigor used to validate the reported protein identifications. A comparison with previously reported identifications (39) was also performed to further validate the results reported here.

Many protein spots yielded more than one confident protein identification; thus, a total of 2570 proteins were identified in the 1367 protein spots indicating co-migration of multiple proteins in the first and second dimensions. For example, spot 1424 contained peptides that confidently identified both a safener protein, In2-1 (51), and a chalcone-flavone isomer-

ase. Both proteins have similar calculated molecular masses (27,029 and 25,601 kDa) and pI values (5.63 and 5.58), making them difficult to resolve through the 2-DE conditions used in this work. However, it cannot be ruled out that the multiple protein identifications in some cases are due to streaky protein spots or due to diffusion of protein spots in close proximity. The presence of co-migrating proteins will complicate quantitative comparison in future experiments and will necessitate alternative methods to confirm quantitative differences (26, 52–57). Of the 2570 proteins identified, the total number of unique proteins represented was 907 (*i.e.* independent accession numbers). The difference between the number of unique sequences (907) and the total number of proteins identified (2570) is attributed to multiple post-translational modifications for each unique sequence or to different protein isoforms (58).

As expected, the vast majority of proteins identified in this

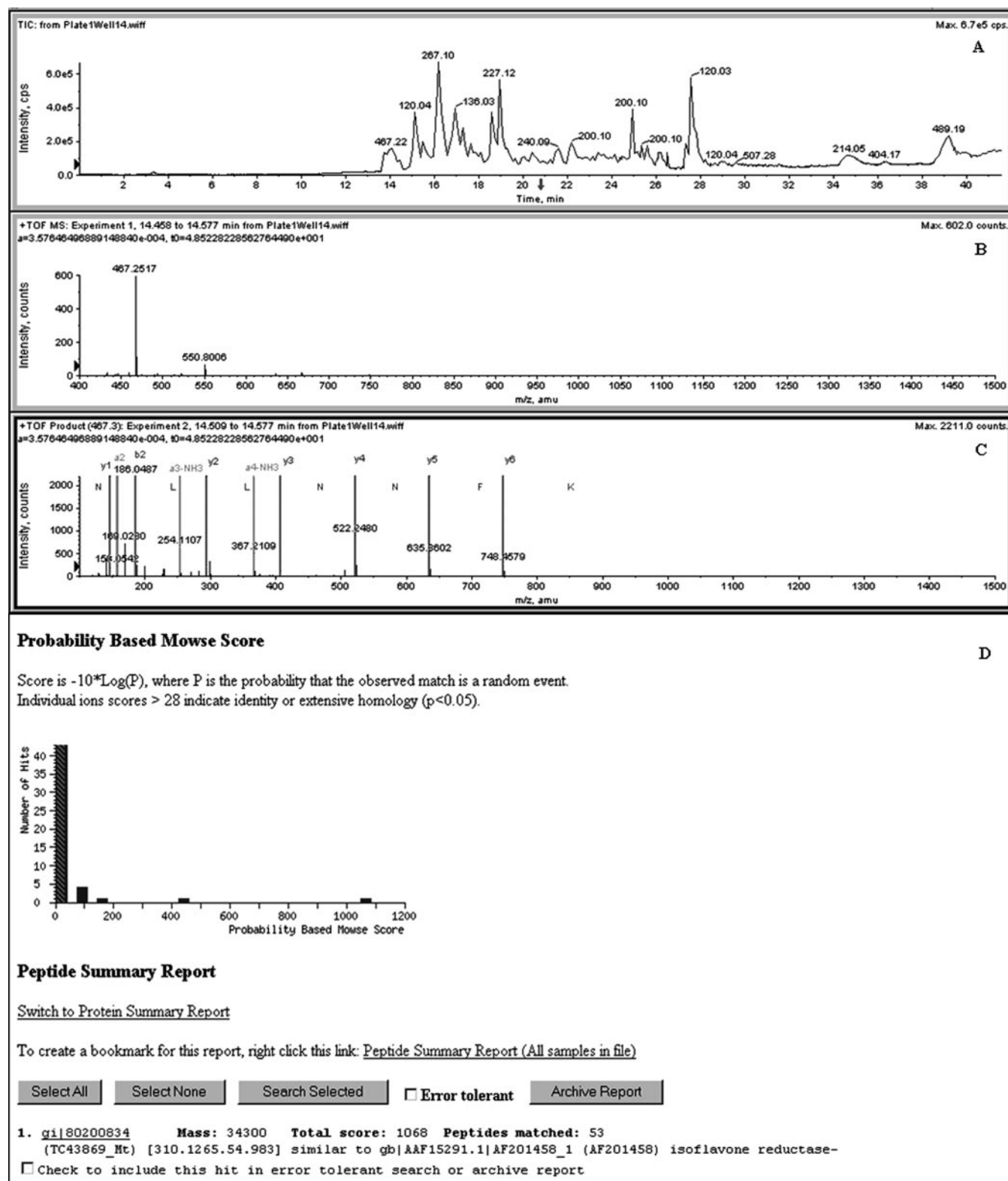


FIG. 2. Representative LC/MS/MS data obtained on an ABI QSTAR Pulsar *i* for spot 1064. A, total ion chromatogram (TIC). B, an example LC/Q-TOF/MS spectra of peptides eluting from the column from 14.458 to 14.577 min. C, tandem LC/Q-TOF/MS/MS mass spectra and experimentally determined sequence for peptide sequence observed at  $m/z$  467.25. D, database search result.

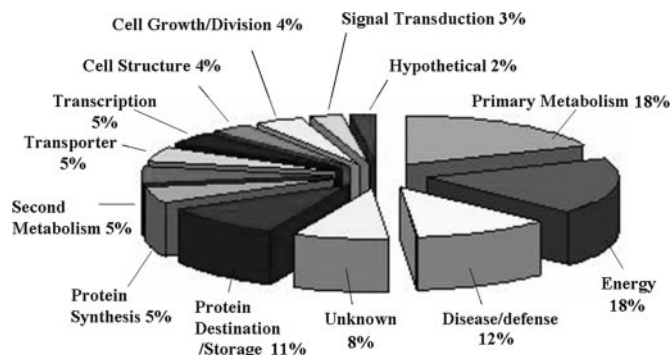


FIG. 3. Distribution of identified protein classes as determined using the NCBI protein-protein BLAST blast with conserved domain search ([www.ncbi.nih.gov/blast/](http://www.ncbi.nih.gov/blast/)), protein families database ([www.sanger.ac.uk/Software/Pfam/](http://www.sanger.ac.uk/Software/Pfam/)), and literature information.

work were soluble proteins because 2-DE analysis of hydrophobic/membrane proteins still remains a significant challenge (59). However, several membrane-associated proteins were identified. These proteins include outer membrane lipoprotein (spot 1519), plasma membrane intrinsic polypeptide (spot 1208), multiple porins (spots 56, 168 and 1500), mitochondrial import receptor subunit TOM20 (spot 1562), several subunits of ATPase (spots 32, 609, and 754), and  $H^+$ -transporting ATPase (spot 1191). Although detection of ATPase subunits has been reported previously in *M. truncatula* (39, 50), other membrane-associated proteins were detected in this species for the first time.

**Functional Classification**—Putative functional annotations for the identified proteins were assigned based on NCBI protein-protein BLAST results obtained with an enabled conserved domain search option, protein families database information ([www.sanger.ac.uk/Software/Pfam/](http://www.sanger.ac.uk/Software/Pfam/)) (60), and/or available literature information. Protein functions were classified into the following categories: primary metabolism, secondary metabolism, energy production, disease/defense responses, signal transduction, transcription, cell growth/division, cell structure, protein destination/storage, protein synthesis, and transport as described previously (61). Fig. 3 illustrates the distribution of proteins in *M. truncatula* cell cultures based on their tentatively assigned functions. Similar to previous reports (39, 50), the major functional categories were metabolism, energy production, disease/defense response, and protein destination/storage. These and other specific functional categories that are expected to be relevant to our long term objectives of comparative proteomics of biotic and abiotic stress responses of *M. truncatula* are discussed below.

**Primary Metabolism and Energy Production**—Proteins involved in primary metabolism are essential to cell growth and maintenance. They accounted for 18% of the identified proteins in the proteome of *M. truncatula* suspension cell cultures. This protein class is of specific interest because significant metabolic reprogramming occurs following exposure of

*M. truncatula* cell cultures to biotic or abiotic stress with carbon being repartitioned from primary metabolism into potentially defensive secondary metabolism (42). Primary metabolic proteins identified in the present study are involved in nitrogen, sulfur, and phosphate metabolism. Examples include an inorganic pyrophosphatase-related protein (spot 1402) that is involved in phosphate assimilation and sulfate adenylyltransferase (spot 668) and adenosine-phosphosulfate kinase (spot 1540) involved in sulfate assimilation (62, 63). Amino acid metabolism-related proteins appeared to be the most abundant, making up about 23% of the identified proteins involved in primary metabolism. Examples include glutamate dehydrogenase 1 (spot 799), aspartate aminotransferase 1 (spot 776), *S*-adenosyl-L-methionine synthetase (spot 639, and note that this enzyme serves as a methyl donor for secondary metabolism), alanine aminotransferase (spot 453),  $\beta$ -cyanoalanine synthase (spot 692, and note this enzyme relative to non-protein amino acid biosynthesis), glutamine synthetase leaf isozyme (spot 755), cysteine synthase (spot 1038), arginase (spot 908), tryptophan synthase (spot 1022), isovaleryl-CoA dehydrogenase (spot 774), phosphoribosyl-ATP pyrophosphohydrolase (At-IE, spot 1264), and ferredoxin-nitrite reductase (spot 282). The conversion of shikimate to chorismate by shikimate kinase (spot 1484), 5-enolpyruvylshikimate-3-phosphate synthase (spot 553), and chorismate synthase (spot 738) leads to the synthesis of phenylalanine, the precursor for the phenylpropanoid secondary metabolites characteristic of the cultures.

Galactokinase (spot 412), glyoxalases I (spot 1223) and II (spot 1274), glycosyl hydrolase family 38 (spot 186), pfkB type carbohydrate kinase protein family (spot 902), fructokinase (spot 1012), and xylulose kinase (spot 38) were among the identified proteins involved in carbohydrate metabolism. Galactokinase is one of the four enzymes in the Leloir pathway that converts galactose to glucose 1-phosphate (64), a more metabolically useful intermediate that can be further converted to glucose 6-phosphate and enter glycolysis for energy production.

Proteins involved in energy production accounted for 18% of all identified proteins. These proteins were dominated by enzymes of glycolysis and the tricarboxylic acid cycle. All of the enzymes involved in glycolysis as well as the tricarboxylic acid cycle were identified except, surprisingly, hexokinase. Examples include fructokinase (spot 1012), aconitase (spot 88), enolase (spot 508), 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (spot 286), phosphoglycerate kinase (spot 780), and triose-phosphate isomerase (spot 1384). These proteins represent some of those identified with the highest confidence.

In addition to glycolytic and tricarboxylic acid cycle enzymes, several proteins involved in the oxidative pentose pathway (D-ribulose-5-phosphate 3-epimerase (spot 1411) and transketolase (spot 171)) leading to the formation of erythrose 4-phosphate, a precursor of the shikimic acid pathway



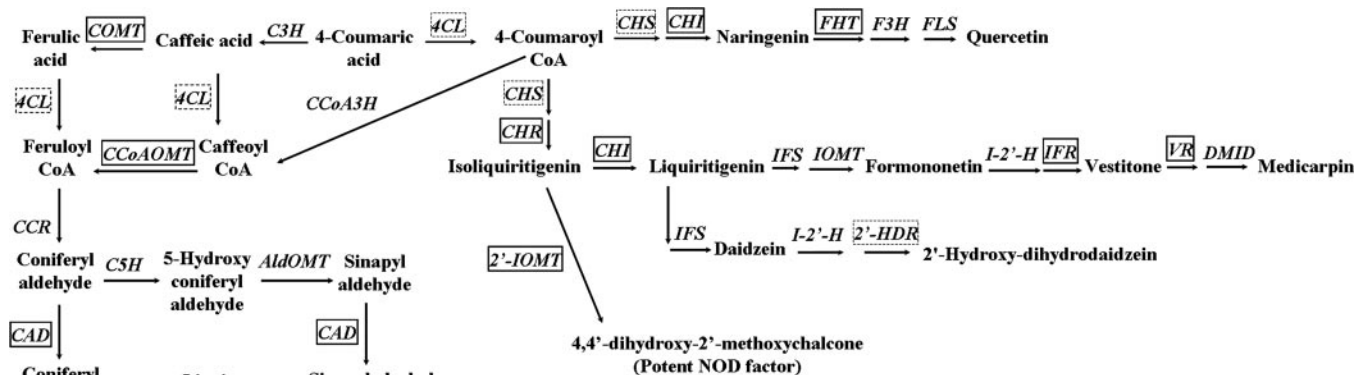


FIG. 4. **Flavonoid and lignin pathways.** Enzymes boxed within solid lines were confidently identified within the *M. truncatula* cell suspension proteome. Often multiple proteins were confidentially identified within the same protein spot, and enzymes boxed within dashed lines were identified as co-eluting with other proteins. 4CL, 4-coumarate:coenzyme A ligase; AldOMT, 5-hydroxyconiferaldehyde-3-O-methyltransferase; C3H, 4-coumarate-3-hydroxylase; C5H, coniferaldehyde-5-hydroxylase; CCoAOMT, caffeoyl-CoA 3-O-methyltransferase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl-alcohol dehydrogenase; CHR, chalcone reductase; CHS, chalcone synthase; COMT, caffeic acid 3-O-methyltransferase; DMID, dihydroxymethyl isoflavone dehydratase; FHT, flavone hydroxytransferase; F3H, flavanone-3-hydroxylase; FLS, flavone synthase; IFR, isoflavone reductase; IFS, isoflavone synthase; I-2'-H, isoflavone-2'-hydroxylase; IOMT, isoflavone-O-methyltransferase; VR, vestitone reductase.

of aromatic amino acid biosynthesis, were identified. The oxidative pentose phosphate pathway is also a major source of reductant (NADPH) for secondary metabolism (65).

**Secondary Metabolism**—Proteins involved in secondary metabolism or natural product biosynthesis constituted 5% of the identified proteins and are of particular interest because of the role of these compounds in plant defense, forage quality, and their pharmacological properties associated with human and animal health (2, 66). Secondary metabolic proteins identified in this study were predominantly associated with flavonoid and lignin biosynthesis (Fig. 4). A number of secondary metabolic enzymes were identified. Although the coverage of the flavonoid and lignin biosynthetic pathways were not as complete as those of glycolysis and the tricarboxylic acid cycle, identification of these enzymes was exciting as many have not been identified previously using a 2-DE/mass spectrometry-based proteomic approach, and the identifications confirm our ability to monitor these important pathways using 2-DE and crude cell lysates. Proteins identified included vestitone reductase (spot 919), isoflavone reductase (spot 1051), chalcone-flavone isomerase (CHI; spot 1052), chalcone reductase (spots 1057 and 1058), caffeic acid 3-O-methyltransferase (spot 823), caffeoyl-CoA O-methyltransferase (spot 1285), and cinnamyl-alcohol dehydrogenase (spot 745). In addition to isoflavone reductase, chalcone isomerase, and vestitone reductase previously identified in *M. truncatula* (39, 50), we also detected other flavonoid-related enzymes, including a putative 2'-hydroxydihydrodaidzein reductase (spot 919) and flavanone 3- $\beta$ -hydroxylase (spot 806). Interestingly most of the flavonoid-related enzymes identified in this work were detected in multiple spots. For example, isoflavone reductase was found in four different spots (spots 991, 1003, 1007, and 1085), chalcone reductase was found in three spots

(spots 1053, 1058, and 1070), and CHI was found in five spots (spots 1424, 1481, 943, 1052, and 1477). The multiple observations of these proteins represent various isoforms that will be useful in evaluating the roles of multigene family members (67). For example, multiple sequence alignment of the chalcone-flavone isomerase sequences indicates that spot 943 contains a Type I CHI, which converts 6'-hydroxychalcone (naringenin chalcone) to 5-hydroxyflavone (naringenin) (Fig. 5). Three spots (spots 1052, 1481, and 1477) contained Type II CHI, which converts 6'-hydroxychalcone as well as converting 6'-deoxychalcone (isoliquiritigenin) to 5-deoxychalcone (liquiritigenin). The CHI identified in spot 1424 shares sequence similarity to both Type I and Type II. The CHI in spot 1424 shares 62% sequence similarity to the Type II CHI sequences and 49% similarity to the Type I sequence. In addition to the global sequence analysis, the sequence from spot 1424 only contains seven of the 29 conserved Type I CHI residues described in Shimada *et al.* (68). Metabolic profiling reveals induction of chalcone derivatives in elicited *M. truncatula* cell suspensions (69), but it has yet to be shown whether specific CHI isoforms are involved in the formation of specific chalcones and their further derivatives. The presence of different CHI isoforms was recently demonstrated in the *M. truncatula* root proteome (50). Proteins involved in flavonoid biosynthesis are often encoded by multiple genes (68, 70), reflecting the complex regulation of this class of molecules that plays such an important role in legume biology. Indeed about 95% of the identified isoflavonoids occur in legumes, and flavonoids have been known to act as inducers of rhizobial *nod* genes in host-specific symbiotic nitrogen fixation (71, 72).

**Defense/Stress-related Proteins**—Proteins involved in defense against biotic and abiotic stress accounted for 12% of

Spot #1481	---MAASITA	ITVENLEYPA	VVTSPVTGKS	YFLGGAGERG	LTIEGNFIKF
Spot #1052	---MAASITA	ITVENLEYPA	VVTSPVTGKS	YFLGGAGERG	LTIEGNFIKF
Spot #1477	---MAASITA	ITVENLEYPA	VVTSPVTGKS	YFLGGAGERG	LTIEGNFIKF
Spot #1424	MATAAPTITG	VKVENIEFPA	VVTTPASPKS	YFLGGAGVRG	LDIDGEFVKF
Spot #943	-HEALPSVTA	LEIENYAFPP	TVKPPGSTNN	FFLGGAGERG	IQIQDKFVKF
	*	**▲ ▲	* ▲* ▲	***** **	* * **
Spot #1481	TAIGVYLEDI	AVASLAACKWK	GKTSQELLDI	LDIFYRDIISG	PFEKLIRGSK
Spot #1052	TAIGVYLEDI	AVASLAACKWK	GKTSQELLDI	LDIFYRDIISG	PFEKLIRGSK
Spot #1477	TAIGVYLEDI	AVASLAACKWK	GKTSQELLDI	LDIFYRDIISG	PFEKLIRGSK
Spot #1424	TGIGIYLEEK	AVASLTPKWK	GKTPSQLFES	LEFYRDIKIG	PFEKFIIRSTK
Spot #943	TAIGVYLDI	AVPYLAEKWK	ARSAHELTDI	VPFFRDIIVT	PFEKFMRTM
	* ** *	**▲ ***	*	▲ * **▲▲*	***** ▲▲
Spot #1481	IRELSGPEYS	RKVMENCVAH	LKSVGTYGDA	EAEAMQKFAE	AFKPINFPPG
Spot #1052	IRELSGPEYS	RKVMENCVAH	LKSVGTYGDA	EAEAMQKFAE	AFKPINFPPG
Spot #1477	IRELSGPEYS	RKVMENCVAH	LKSVGTYGDA	EAEAMQKFAE	AFKPINFPPG
Spot #1424	VRTLEGSEYV	RKVSENCIAH	MKSEGTYGDA	EKAIQEFRE	AFKDQFFPPG
Spot #943	ILPLTGHQYS	EKVSENCVAI	WKSIGIYTDE	EAKAIDKFVS	VFKDETFFPG
	▲▲* ▲*	** *** *	** ▲▲▲*	* *	▲▲▲ ****
Spot #1481	ASVFYRQSPD	--GILGLSFS	PDTSIPEKEA	ALIENKAVSS	AVLETMIGE
Spot #1052	ASVFYRQSPD	--GILGLSFS	PDTSIPEKEA	ALIENKAVSS	AVLETMIGE
Spot #1477	ASVFYRQSPD	--GILGLSFS	PDTSIPEKEA	ALIENKAVSS	AVLETMIGE
Spot #1424	TAAFYRQSPN	--GALGIRFS	KDETIPEHEY	AVINNKLPLSE	AVLETMIGE
Spot #943	SSILFTVSPK	GLGSITISFS	KDGSIPVET	AVIENKLLSQ	AVLESMIGAH
	▲▲▲▲	* * **	▲▲▲ *** *	* * **▲ *	*****▲▲
Spot #1481	AVSPDLKRCL	AARLPALLNE	GAFKIGN	} Type II	
Spot #1052	AVSPDLKRCL	AARLPALLNE	GAFKIGN		
Spot #1477	AVSPDLKRCL	AARLPALLNE	GAFKIGN		
Spot #1424	PVSPALKESL	ATRFYEFLKI	DNFNIRN	} Type I	
Spot #943	GVSPAAKQSL	ASRLSKLFKE	GG-NANN		
	▲▲▲▲ ▲* ▲*	* * *	*		

FIG. 5. Sequence alignment of the chalcone-flavanone isomerase isoforms. Shaded areas represent peptides identified in the HPLC/Q-TOF/MS/MS. Asterisks show identical residues. Vertical arrows indicate residues that are conserved in Type I chalcone-flavanone isomerases.

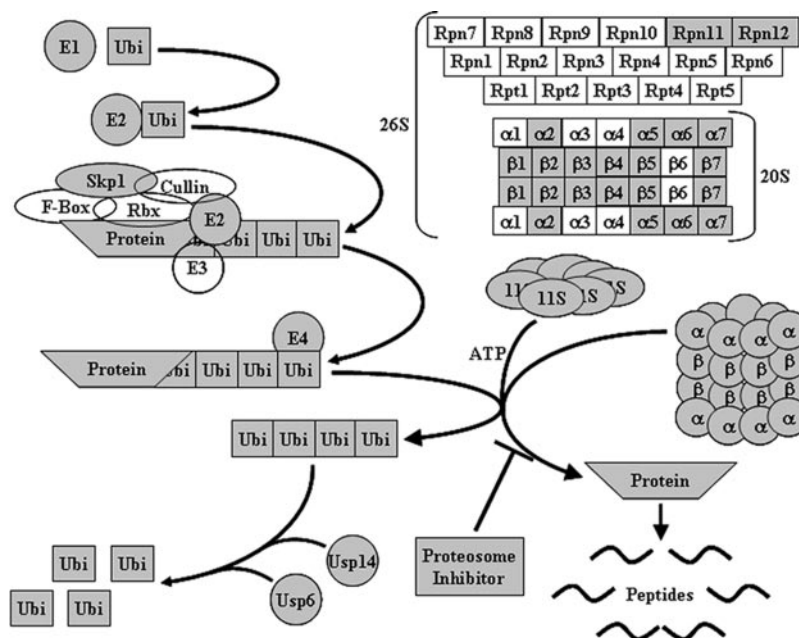
all the identified proteins. These proteins protect plants against pathogenic fungi, bacteria, and viruses and adverse environmental conditions and can be induced by a variety of biotic and abiotic stimuli such as wounding, pathogen infection, or environmental stresses. The most abundant protein in the 2-DE proteome reference map was identified as an abscisic acid-responsive protein (ABR-17, spot 1644) with a high MOWSE score (783) and protein sequence coverage (98%). This protein represents the highest sequence coverage achieved in this work, and peptides missed in this identification were below the TOF-MS lower  $m/z$  cutoff (scan range  $m/z$ , 400–1500) used in the analyses. Although the photosynthetic carbon fixation enzyme ribulose-bisphosphate carboxylase is the most dominant protein in leaf and stem tissues, ABR-17 is the most abundant protein found in roots and root-derived cell cultures (39, 50). Although the exact function of ABR-17 is still unclear, it has high sequence similarity to intracellular PR proteins and to stress response-related proteins (73). It is interesting that the five most abundant proteins in the cultures, ABR-17 (spot 1644), thaumatin-like protein PR-5b (spot 1491), 1,3- $\beta$ -glucanase (spot 1103), calmodulin

(spot 1651) and In2-1 protein (spot 1424) are all PR- or stress-related proteins. Although calmodulin is generally thought to be involved in signal transduction, some isoforms are induced by fungal elicitors and pathogens, and transgenic plants expressing these proteins show enhanced resistance to a wide range of bacteria, fungi, and virus (74). The most abundant proteins in *M. truncatula* roots are common metabolic enzymes and transport proteins (50). The preponderance of disease/stress-related proteins in *M. truncatula* suspension cell cultures may be attributed to the stress associated with growth of the cells in culture. Indeed some PR-related proteins have been demonstrated to be induced by the culture process (75).

Additional classes of defense/stress-related proteins identified in this work included proteinase inhibitors, peroxidases, proteinases, and carbohydrases. Although a significant number of the identified disease/stress-related proteins have been reported previously (37, 39, 50), many unique *M. truncatula* proteins were revealed for the first time. For example, spot 454 was found to contain peptides consistent with a *G. max* matrix metalloproteinase, GmMMP2 (accession number



FIG. 6. Ubiquitination and proteolysis is initiated through activation of the Skp1, cullin, F-box (SCF) complex that also includes a RING finger protein (Rbx) by the COP9 signalosome. Enzymes and proteins associated with this processes and identified in this work are listed in gray boxes. Ubi, ubiquitin; Usp6, ubiquitin-specific protease 6; Usp14, ubiquitin-specific protease 14; Rpn1–12, 26 S regulatory subunits; Rpt1–5, 26 S regulatory subunits;  $\alpha$ 1–7, 20 S structural subunits;  $\beta$ 1–7, 20 S catalytic subunits; 11S, 11 S proteasome activator subunits.



AAL27029.1). This is the first report of a matrix metalloproteinase from *M. truncatula*. The soybean metalloproteinase gene, *GmMMP2*, is single copy (76). However, three different spots (spots 389, 454, and 484) with close pI values and molecular weights were found to contain peptides that can be associated with GmMMP. This suggests the presence of multiple matrix metalloproteinase (MMP) isoforms in *M. truncatula* (58), which is further supported by multiple tentative consensus sequences for MMPs in The Institute for Genomic Research *M. truncatula* gene indices ([www.tigr.org/tigr-scripts/tgi/T\\_index.cgi?species=medicago](http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=medicago)). GmMMP2 is a zinc-binding protein and is thought to play a role in pathogen defense (76). Unlike many other defense-related proteins, the activation of MMP2 is not dependent on salicylic acid or jasmonic acid signaling pathways (76). To date, only three plant MMP proteins from soybean leaves and buckwheat seeds (76–78) and two cDNAs encoding MMPs from *Arabidopsis* and cucumber have been identified (79, 80). Other disease/stress-related proteins that were not identified in the previous proteomic studies of *M. truncatula* include haloacid dehalogenase-like hydrolase family (spot 1376), which shares sequence similarity with sucrose-6-phosphate phosphohydrolase and a sequence from *Arabidopsis* that has a haloacid dehalogenase-like domain, osmotin-like protein precursor (spot 1412), UVB resistance protein-like protein (spot 705), thaumatin-like protein PR-5b (spot 3), and MtN13 (spot 1627). MtN13 is a protein identified in *M. truncatula* that is closely related to the PR-10 family (81). Unlike some other members of the PR-10 family found in *M. truncatula* such as MtPR10-1, which is constitutively expressed in roots and pathogen-inducible in leaves, MtN13 was reported to be exclusively expressed in roots during nodulation and was found specifically in the nodule outer cortex (81). However, recent work revealed

high expression of MtN13 transcripts in alfalfa trichomes (82). MtN13 expression may be promoted by the culture process as found with other PR-related proteins (75) or its expression in the cultures may reflect a “memory” effect; cell cultures tend to retain portions of the gene expression pattern of the tissues from which they were derived (41).

The protein 12-oxophytodienoate reductase (spot 799) is involved in jasmonic acid biosynthesis (for a review, see Ref. 83). Jasmonic acid signaling is typically induced as a defense response to wounding such as herbivore damage or chewing insects. Among other proteins induced during the wound response and jasmonate signaling are the COP9 signalosome and 26 S proteasome complex proteins. These proteins are critical in controlled proteolysis (18, 84), which represents an alternate mechanism of regulating protein activity and one that cannot be evaluated at the transcript level.

**Transport**—Transport proteins accounted for 5% of the identified proteins and included proteins putatively involved in membrane trafficking, intracellular trafficking, and nutrient transport. Examples include an ATP-binding cassette transporter family protein (spot 1272), chaperonin (CPN60/HSP60, spot 304), caseinolytic protease subclass B (spot 84), dnaK-type molecular chaperone binding protein isoform B (spot 403), ferritin (spot 665), GTP-binding proteins (spot 712), vacuolar sorting receptor AtELP2b (spot 141), transporter glucose regulated protein E-like protein (spot 401), probable potassium channel  $\beta$  chain KB1 (spot 181), and outer plastidial membrane protein porin (spot 1193). Most of the proteins in this category have not been identified in previous reports of the *M. truncatula* proteome (39, 50). Some of these proteins may have multiple functions. For example, several GTP-binding proteins with different accession numbers were identified, reflecting multiple groups of the superfamily, which is divided

into five groups: Ras/Ras-like, Rho/Rac, Ypt/Rab, Ran/TC4, and Arf/Sar. GTP-binding protein (spot 885, accession number T06448) belongs to the Ras subfamily (85), whereas another GTP-binding protein (spot 720, accession number AAF65513.1) was related to Ran/TCR4 group using NCBI BLAST and Pfam domain searches. These proteins are implicated in vesicular transport, cell proliferation, signal transduction, and cytoskeletal organization (86). Transport of some natural products is believed to occur via vesicular transport (87), but this area is still in its infancy.

**Protein Destination and Storage (Proteolysis)**—This class includes proteolysis-related proteins, and a large number of these proteins were identified in this study. Proteolysis is initiated by the activation of the Skp1 (spot 1549), cullin, F-box (SCF) complex that also includes a RING finger protein (Rbx) by the COP9 signalosome downstream of jasmonate signaling (Fig. 6). A large number of proteins from the 26 S proteasome as well as the enzymes involved in protein ubiquitination were identified and represent 11% of the total proteins identified. Clearly the cell cultures provide a good system to investigate the contribution of controlled proteolysis to various stress responses at the protein level. The majority of the 20 S catalytic complex proteins (*i.e.* 11 of 14 total) were identified except one  $\beta$  subunit and three  $\alpha$  subunits, whereas the 19 S regulatory complex was only represented by Rpn11 and Rpn12 (Fig. 6). In addition, both the proteasome 11 S activator complex and a proteasome inhibitor were identified. Each of the ubiquitin ligase enzymes (E1, E2, and E4) and a ubiquitin fusion degradation protein with putative E3 ligase-like activity were present within the proteome. Finally two ubiquitin-specific proteases (Usp6 and Usp14) responsible for the degradation of polyubiquitin complete the cycle. The detection of a nearly complete proteasome degradation pathway within the *M. truncatula* cell suspension proteome makes the cultures a novel system for studying both upstream signaling processes such as jasmonic acid induction of the proteasome (88) as well as the regulation of downstream controlled proteolysis of targeted proteins.

**Other Identified Protein Functions**—Other identified protein categories included transcription (5%), protein synthesis (5%), cell structure (4%), cell growth/division (4%), signal transduction (3%), unknown (8%), and hypothetical proteins (2%). Signal transduction-related proteins identified include mitogen-activated protein kinase kinase kinase-like protein kinase (spot 389), calmodulin (spot 1651) (see above), cystathionine  $\beta$  synthase domain-containing protein (spot 1593), transforming growth factor- $\beta$  receptor-interacting protein (spot 975), Ran GTPase activase (spot 316), Ara4-interacting protein (spot 66), 14-3-3-like protein (spot 1243), Rho GDP dissociation inhibitor (spot 1234), and others (see Supplemental Table 1). Proteins involved in protein synthesis were mainly 30, 40, and 60 S ribosomal proteins, elongation factors (spot 490), initiation factors (spot 1626), and tRNA synthetases (spot 456) and ligases (spot 819). Approximately 120 transcrip-

tion-related proteins were identified and included DNA-binding proteins, RNA-binding proteins, RNA recognition motif-containing protein, RSZp22 splicing factor, transcription factor BTF3-related, and the largest subunit of DNA-directed RNA polymerase II.

Several glycosyl hydrolase family proteins, pectinesterase precursor, and fasciclin-like arabinogalactan-protein (FLA10) are among those identified to be involved in cell structure. They are mainly involved in the hydrolysis of cell wall polysaccharides, leading to the modification of cell wall structures. Examples of cell growth/division-related proteins were actin-like proteins, actin-depolymerizing factor 2, prohibitin, tubulin-related proteins, CIG2, DEAD/DEAH box helicase, proliferating cell nuclear antigen, and KAP-2. Although only 4% of the identified proteins, proteins involved in cell growth/division are critical for normal cell growth. KAP-2 is an interesting protein found in bean (*Phaseolus vulgaris*) and *M. truncatula* that shares sequence similarity to the large subunit of mammalian Ku autoantigen (89), a protein demonstrated to be involved in DNA damage repair (90). However, it was also demonstrated to be a transcription factor that binds to the H-box in a bean chalcone synthase promoter (89). The presence of KAP-2 is not surprising as multiple chalcone biosynthetic enzymes were identified in this study, and chalcone synthase transcripts are induced in cultures in response to yeast elicitor (69).

**Identified Root-specific Proteins Suggest Cell Culture Memory Effects**—Root specific proteins identified included root border cell-specific protein (spot 1183), non-symbiotic hemoglobin (spot 1635), nodulin (spot 478), and an extracellular dermal glycoprotein precursor (spot 570). Plant root border cells serve as a chemical, physical, and biological interface between roots and soil to protect roots from fungal pathogens and to detoxify environmental pollutants (91, 92). Although its function is still unknown, the root border-specific protein was found to contain domains similar to heme iron utilization proteins using NCBI protein BLAST. This suggests a possible role of this root border-specific protein in inorganic ion transport and metabolism. Although non-symbiotic hemoglobin can be found in various tissues, the particular protein encoded by the *Mhb1* gene has been reported as root-specific and induced by hypoxia (93), a phenomenon normally occurring in suspension cells in media or roots in the soil. The nodulin identified in this work is a novel 53-kDa nodulin of the symbiosome membrane originally identified in soybean (protein accession number AAC72337.1). Its exact function is still unknown; however, a conserved domain search using NCBI BLAST reveals its similarity to flotillins, suggesting that it is involved in intracellular trafficking.

Other root-specific proteins were also identified. Inosine-monophosphate dehydrogenase (spot 425) catalyzes the rate-limiting step in *de novo* biosynthesis of purine and is a postulated key enzyme in nitrogen assimilation in ureide-exporting nodules (94); RNA gel blot analysis of soybean

inosine-monophosphate dehydrogenase revealed its expression to be nodule-specific. Pyridoxal kinase-like protein SOS4 (spot 1151) is an enzyme involved in the biosynthesis of pyridoxal 5-phosphate. Pyridoxal 5-phosphate, an active form of vitamin B<sub>6</sub>, is an important cofactor for various enzymes, acts as a ligand for certain ion transporters (95), and is required for root hair development (96). The presence of these apparently root-specific proteins in the cell suspensions, which were derived from roots but only after multiple passages as callus and liquid suspension, are consistent with the concept of "culture memory."

**Conclusions**—A high resolution 2-DE proteome reference map for *M. truncatula* was generated. Of the 1661 proteins visualized by 2-DE, 1367 were identified with high confidence using nano-HPLC/Q-TOF/MS/MS. The identified proteins were functionally classified, and specific classes related to stress and secondary metabolism were discussed. All enzymes involved in the tricarboxylic acid cycle were observed as well as the majority of the glycolytic enzymes. However, in this sea of primary metabolic enzymes, a large number of enzymes (*i.e.* 18) involved in secondary metabolism were also observed. The observation of a significant number and multiple isoforms of many secondary metabolic enzymes further suggests exciting opportunities for the study of secondary metabolism using a comparative 2-DE proteomic approach. Similarly the sizable number of proteins classified as defense/stress-related proteins will enable the evaluation of stress at the proteome level. The copious number of enzymes identified in the ubiquitination/proteasome pathway will allow visualization of controlled proteolysis in future experiments. Interestingly many of the identified proteins were root-associated or root-specific. This suggests that there may be a memory effect associated with the cell culture process as the cultures were initiated from root tissue. The memory effect further suggests that information gained from cell culture experiments should be representative and relevant to root physiology.

This work represents the most extensive investigation of the *M. truncatula* proteome to date, and the reference map provides a starting point for ongoing functional genomic studies associated with biotic/abiotic stress and natural product biosynthesis in *M. truncatula*. These studies include comparative proteomics as well as transcriptomics and metabolomics. Many of the genes and gene families represented in the identified proteins and isozymes reported in this study can also be evaluated at the transcript level using microarrays; however, the proteome context allows visualization of specific processes such as post-translational modifications and activation of enzymes that are transparent using transcriptome analyses only. Although the current depth of coverage of the proteome is still significantly less than that achievable with genome-based microarrays, we fully believe that integration of the protein context with the transcriptome and even metabolome data will provide the most comprehensive and

informative view of the system response and allow us to better establish casual relationships between the transcriptome, proteome, and metabolome. Currently these large scale biochemical profiling technologies are leading to new discoveries and understanding of *Medicago* biology (39, 42, 69, 82, 97–101), and we envision that they will continue to provide even greater insight.

**Acknowledgments**—We thank Drs. Kirankumar S. Mysore and Richard S. Nelson for critical reading of this manuscript, Lianjiang Wang for generating the custom protein database, and Alisha Raney for clerical assistance.

\* This work was supported by National Science Foundation Grant DBI-0109732 and by The Samuel Roberts Noble Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

¶ To whom correspondence should be addressed: Plant Biology Division, The Samuel Roberts Noble Foundation, 2510 Sam Noble Pkwy., Ardmore, OK 73402. Fax: 580-224-6692; E-mail: [lwsumner@noble.org](mailto:lwsumner@noble.org).

## REFERENCES

- Dixon, R. A., and Steele, C. L. (1999) Flavonoids and isoflavonoids—a gold mine for metabolic engineering. *Trends Plant Sci.* **4**, 394–400
- Dixon, R. A., and Sumner, L. W. (2003) Legume natural products: understanding and manipulating complex pathways for human and animal health. *Plant. Physiol.* **131**, 878–885
- Sarkar, F. H., and Li, Y. (2003) Soy isoflavones and cancer prevention. *Cancer Investig.* **21**, 744–757
- United States Department of Agriculture-National Agricultural Statistics Service (2004) *Agricultural Statistics 2004*, United States Government Printing Office, Washington, D. C.
- Barker, D. G., Bianchi, S., Blondon, F., Datteé, Y., Duc, G., Essad, S., Flament, P., Gallusci, P., Génier, G., Guy, P., Muel, X., Tourneur, J., Dénarié, J., and Huguet, T. (1990) *Medicago truncatula*, a model plant for studying the molecular genetics of the Rhizobium-legume symbiosis. *Plant Mol. Biol. Rep.* **8**, 40–49
- Bell, C. A., Dixon, R. A., Farmer, A. D., Flores, R., Inman, J., Gonzales, R. A., Harrison, M. J., Paiva, N. L., Scott, A. D., Weller, J. W., and May, G. D. (2000) The *Medicago* genome initiative: a model legume database. *Nucleic Acid Res.* **29**, 1–4
- Trieu, A. T., Burleigh, S. H., Kardailsky, I. V., Maldonado-Mendoza, I. E., Versaw, W. K., Blaylock, L. A., Shin, H., Chiou, T. J., Katagi, H., Dewbre, G. R., Weigel, D., and Harrison, M. J. (2000) Transformation of *Medicago truncatula* via infiltration of seedlings or flowering plants with *Agrobacterium*. *Plant J.* **22**, 531–541
- Cook, D. R. (1999) *Medicago truncatula*—a model in the making! *Curr. Opin. Plant Biol.* **2**, 301–304
- Cook, D. R., VandenBosch, K., de Bruijn, F. J., and Huguet, T. (1997) Model legumes get the nod. *Plant Cell* **9**, 275–281
- Thoquet, P., Gherardi, M., Journet, E., Kereszt, A., Ane, J., Prosperi, J., and Huguet, T. (2002) The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes. *BMC Plant Biol.* **2**, 1–13
- Choi, H.-K., Kim, D., Uhm, T., Limpens, E., Lim, H., Mun, J.-H., Kalo, P., Penmetsa, R. V., Seres, A., Kulikova, O., Roe, B. A., Bisseling, T., Kiss, G. B., and Cook, D. R. (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. *Genetics* **166**, 1463–1502
- May, G. D., and Dixon, R. A. (2004) *Medicago truncatula*. *Curr. Biol.* **9**, 180–181



13. Chory, J., Ecker, J. R., Briggs, S., Caboche, M., Coruzzi, G. M., Cook, D., Dangl, J., Grant, S., Gueriot, M. L., Henikoff, S., Martienssen, R., Okada, K., Raikhel, N. V., Somerville, C. R., and Weigel, D. (2000) National Science Foundation-Sponsored Workshop Report: "The 2010 project" functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. *Plant Physiol.* **123**, 423–426
14. Somerville, C., and Somerville, S. (1999) Plant functional genomics. *Science* **285**, 380–383
15. Aharoni, A., and Vorst, O. (2002) DNA microarrays for functional plant genomics. *Plant Mol. Biol.* **48**, 99–118
16. Wu, S. H., Ramonell, K., Gollub, J., and Somerville, S. (2001) Plant gene expression profiling with DNA microarrays. *Plant Physiol. Biochem.* **39**, 917–926
17. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720–1730
18. Serino, G., and Deng, X. (2003) The COP9 signalsome: regulating plant development through the control of proteolysis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **54**, 165–182
19. (2003) Focus: proteomics. *Nat. Biotechnol.* **21**, 213–340
20. (2004) Focus: systems biology. *Nat. Biotechnol.* **22**, 1191–1274
21. (2004) Focus: plant proteomics. *Phytochemistry* **65**, 1239–1863
22. Blackstock, W. P., and Weir, M. P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* **17**, 121–127
23. Görg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., and Weiss, W. (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**, 1037–1053
24. Herbert, B. (1999) Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis* **20**, 660–663
25. Figeys, D. (2003) Proteomics in 2002: a year of technical development and wide-ranging applications. *Anal. Chem.* **75**, 2891–2905
26. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207
27. Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S., and Garrels, J. I. (1999) A sampling of the yeast proteome. *Mol. Cell. Biol.* **19**, 7357–7368
28. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003) Global analysis of protein expression in yeast. *Nature* **425**, 737–741
29. Anderson, N. G., Matheson, A., and Anderson, N. L. (2001) Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics* **1**, 3–12
30. Anderson, N. L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R. S., Conrads, T. P., Veenstra, T. D., Adkins, J. N., Pounds, J. G., Fagan, R., and Lobley, A. (2004) The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **3**, 311–326
31. Houtman, R., Krijgsveld, J., Kool, M., Romijn, E. P., Redegeld, F. A., Nijkamp, F. P., Heck, A. J., and Humphery-Smith, I. (2003) Lung proteome alterations in a mouse model for nonallergic asthma. *Proteomics* **3**, 2008–2018
32. Edvardsson, U., Alexandersson, M., Brockenhuus von Lowenhielm, H., Nystrom, A. C., Ljung, B., Nilsson, F., and Dahllof, B. (1999) A proteome analysis of livers from obese (ob/ob) mice treated with the peroxisome proliferator WY14,643. *Electrophoresis* **20**, 935–942
33. Guerreiro, N., Djordjevic, M. A., and Rolfe, B. G. (1999) Proteome analysis of the model microsymbiont *Sinorhizobium meliloti*: isolation and characterisation of novel proteins. *Electrophoresis* **20**, 818–825
34. Morris, A. C., and Djordjevic, M. A. (2001) Proteome analysis of cultivar-specific interactions between *Rhizobium leguminosarum* biovar trifolii and subterranean clover cultivar Woogenellup. *Electrophoresis* **22**, 586–598
35. van Wijk, K. J. (2001) Challenges and prospects of plant proteomics. *Plant Physiol.* **126**, 501–508
36. Kersten, B., Burkle, L., Kuhn, E. J., Giavalisco, P., Konthur, Z., Lueking, A., Walter, G., Eickhoff, H., and Schneider, U. (2002) Large-scale plant proteomics. *Plant Mol. Biol.* **48**, 133–141
37. Bestel-Corre, G., Dumas-Gaudot, E., Poinso, V., Dieu, M., Dierick, J. F., van Tuinen, T. D., Remacle, J., Gianinazzi-Pearson, V., and Gianinazzi, S. (2002) Proteome analysis and identification of symbiosis-related proteins from *Medicago truncatula* Gaertn. by two-dimensional electrophoresis and mass spectrometry. *Electrophoresis* **23**, 122–137
38. Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R. D., and Burstin, J. (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol.* **133**, 664–682
39. Watson, B. S., Asirvatham, V. S., Wang, L., and Sumner, L. W. (2003) Mapping the proteome of barrel medic (*Medicago truncatula*). *Plant Physiol.* **131**, 1104–1123
40. Imin, N., De Jong, F., Mathesius, U., van Noorden, G., Saeed, N., Wang, X., Rose, R., and Rolfe, B. (2004) Proteome reference maps of *Medicago truncatula* embryogenic cell cultures generated from single protoplasts. *Proteomics* **4**, 1883–1896
41. Dixon, R. A. (1980) Plant tissue culture methods in the study of phytoalexin induction, in *Tissue Culture Methods for Plant Pathologists* (Ingram, D. S., and Helgeson, J. P., eds) p. 185–196, Blackwell Scientific Publications, Oxford
42. Broeckling, C. B., Huhman, D. V., Farag, M. A., Smith, J. T., May, G. D., Mendes, P., Dixon, R. A., and Sumner, L. W. (2005) Metabolic profiling of *Medicago truncatula* cell cultures reveals effects of biotic and abiotic elicitors on metabolism. *J. Exp. Bot.* **56**, 323–336
43. Molloy, M. P., Herbert, B. R., Walsh, B. J., Tyler, M. I., Traini, M., Sanchez, J. C., Hochstrasser, D. F., Williams, K. L., and Gooley, A. A. (1998) Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. *Electrophoresis* **19**, 837–844
44. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254
45. Creasy, D. M., and Cottrell, J. S. (2002) Error tolerant searching of uninterpreted tandem mass spectrometry data. *Proteomics* **2**, 1426–1434
46. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567
47. Molloy, M., Brzezinski, E., Hang, J., McDowell, M., and VanBogelen, R. (2003) Overcoming technical variation and biological variation in quantitative proteomics. *Proteomics* **3**, 1912–1919
48. Asirvatham, V. S., Watson, B. S., and Sumner, L. W. (2002) Analytical and biological variances associated with proteomic studies of *Medicago truncatula* by 2-DE. *Proteomics* **2**, 960–968
49. Candiano, G., Musante, L., Bruschi, M., Ghiggeri, G. M., Herbert, B., Antonucci, F., and Righetti, P. G. (2002) Two-dimensional maps in soft immobilized pH gradient gels: a new approach to the proteome of the third millennium. *Electrophoresis* **23**, 292–297
50. Mathesius, U., Keijzers, G., Natera, S. H., Weinman, J. J., Djordjevic, M. A., and Rolfe, B. G. (2001) Establishment of a root proteome reference map for the model legume *Medicago truncatula* using the expressed sequence tag database for peptide mass fingerprinting. *Proteomics* **1**, 1424–1440
51. Edwards, R., Dixon, D. P., and Walbot, V. (2000) Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health. *Trends Plant Sci.* **5**, 193–198
52. Whitelegge, J. (2004) Mass spectrometry for high throughput quantitative proteomics in plant research: lessons from thylakoid membranes. *Plant Physiol. Biochem.* **42**, 919–927
53. Righetti, P., Castagna, A., Antonucci, F., Piubelli, C., Cecconi, D., Campostrini, N., Antonoli, P., Astner, H., and Hamdan, M. (2004) Critical survey of quantitative proteomics in two-dimensional electrophoretic approaches. *J. Chromatogr. A* **1051**, 3–17
54. Righetti, P., Campostrini, N., Pascali, J., Hamdan, M., and Astner, H. (2004) Quantitative proteomics: a review of different methodologies. *Eur. J. Mass Spectrom. (Chichester, Eng.)* **10**, 335–348
55. Tao, W., and Aebersold, R. (2003) Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. *Curr. Opin. Biotechnol.* **14**, 110–118
56. Ong, S., Foster, L., and Mann, M. (2003) Mass spectrometric-based approaches in quantitative proteomics. *Methods* **29**, 124–130
57. Goshe, M., and Smith, R. (2003) Stable isotope-coded proteomic mass spectrometry. *Curr. Opin. Biotechnol.* **14**, 101–109
58. Claverol, S., Burlet-Schiltz, O., Gairin, J. E., and Monsarrat, B. (2003) Characterization of protein variants and post-translational modifications: ESI-MS<sup>n</sup> analyses of intact proteins eluted from polyacrylamide

- gels. *Mol. Cell. Proteomics* **2**, 483–493
59. Luche, S., Santoni, V., and Rabilloud, T. (2003) Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics* **3**, 249–253
60. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002) The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280
61. Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirkse, W., Van Staveren, M., Stiekema, W., Drost, L., Ridley, P., Hudson, S. A., Patel, K., Murphy, G., Piffanelli, P., Wedler, H., Wedler, E., Wambutt, R., Weitzenegger, T., Pohl, T. M., Terry, N., Gielen, J., Villarreal, R., De Clerck, R., Van Montagu, M., Lecharny, A., Auborg, S., Gy, I., Kreis, M., Lao, N., Kavanagh, T., Hempel, S., Kotter, P., Entian, K. D., Rieger, M., Schaeffer, M., Funk, B., Mueller-Auer, S., Silvey, M., James, R., Montfort, A., Pons, A., Puigdomenech, P., Douka, A., Voukelatou, E., Milioni, D., Hatzopoulos, P., Piravandi, E., Obermaier, B., Hilbert, H., Dusterhoft, A., Moores, T., Jones, J. D., Eneva, T., Palme, K., Benes, V., Rechman, S., Ansoorge, W., Cooke, R., Berger, C., Delsen, M., Voet, M., Volckaert, G., Mewes, H. W., Klosterman, S., Schueller, C., and Chalwatzis, N. (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* **391**, 485–488
62. Fujimori, K., and Ohta, D. (1998) Isolation and characterization of a histidine biosynthetic gene in *Arabidopsis* encoding a polypeptide with two separate domains for phosphoribosyl-ATP pyrophosphohydrolase and phosphoribosyl-AMP cyclohydrolase. *Plant Physiol.* **118**, 275–283
63. Geller, D. H., Henry, J. G., Belch, J., and Schwartz, N. B. (1987) Copurification and characterization of ATP-sulfurylase and adenosine-5'-phosphosulfate kinase from rat chondrosarcoma. *J. Biol. Chem.* **262**, 7374–7382
64. Holden, H. M., Rayment, I., and Thoden, J. B. (2003) Structure and function of enzymes of the Leloir pathway for galactose metabolism. *J. Biol. Chem.* **278**, 43885–43888
65. Pryke, J. A., and Rees, T. (1977) The pentose phosphate pathway as a source of NADPH for lignin synthesis. *Phytochemistry* **16**, 557–560
66. Dixon, R. A. (2001) Natural products and disease resistance. *Nature* **411**, 843–847
67. Sappl, P., Heazlewood, J., and Millar, A. (2004) Untangling multi-gene families in plants by integrating proteomics into functional genomics. *Phytochemistry* **65**, 1517–1530
68. Shimada, N., Aoki, T., Sato, S., Nakamura, Y., Tabata, S., and Ayabe, S. (2003) A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso)flavonoids in *Lotus japonicus*. *Plant Physiol.* **131**, 941–951
69. Suzuki, H., Reddy, M. S., Naoumkina, M., Aziz, N., May, G. D., Huhman, D. V., Sumner, L. W., Blount, J. W., Mendes, P., and Dixon, R. A. (2005) Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures. *Planta* **220**, 696–707
70. Akada, S., and Dube, S. K. (1995) Organization of soybean chalcone synthase gene clusters and characterization of a new member of the family. *Plant Mol. Biol.* **29**, 189–199
71. Hegnauer, R., and Gpuyer-Barkmeijer, R. J. (1993) Relevance of seed polysaccharides and flavonoids for the classification of the leguminosae: a chemotaxonomic approach. *Phytochemistry* **34**, 3–16
72. Aoki, T., Akashi, T., and Ayabe, S. (2000) Flavonoids of leguminous plants: structure, biological activity, and biosynthesis. *J. Plant Res.* **113**, 475–488
73. Iturriaga, E. A., Leech, M. J., Barratt, D. H., and Wang, T. L. (1994) Two ABA-responsive proteins from pea (*Pisum sativum* L.) are closely related to intracellular pathogenesis-related proteins. *Plant Mol. Biol.* **24**, 235–240
74. Heo, W. D., Lee, S. H., Kim, M. C., Kim, J. C., Chung, W. S., Chun, H. J., Lee, K. J., Park, C. Y., Park, H. C., Choi, J. Y., and Cho, M. J. (1999) Involvement of specific calmodulin isoforms in salicylic acid-independent activation of plant disease resistance responses. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 766–771
75. Barratt, D. H. P., Domoney, C., and Wang, T. L. (1989) Purification and partial characterization of two abscisic-acid-responsive proteins induced in cultured embryos of *Pisum sativum* L. *Planta* **180**, 16–23
76. Liu, Y., Dammann, C., and Bhattacharyya, M. K. (2001) The matrix metalloproteinase gene GmMMP2 is activated in response to pathogenic infections in soybean. *Plant Physiol.* **127**, 1788–1797
77. Graham, J. S., Xiong, J., and Gillikin, J. W. (1991) Purification and developmental analysis of a metalloendoproteinase from leaves of *Glycine max*. *Plant Physiol.* **97**, 786–792
78. Belozersky, M., Dunaevsky, Y., and Voskoboynikova, N. (1990) Isolation and properties of a metalloproteinase from buckwheat (*Fagopyrum esculentum*) seeds. *Biochem. J.* **272**, 677–682
79. Delorme, V. G., McCabe, P. F., Kim, D. J., and Leaver, C. J. (2000) A matrix metalloproteinase gene is expressed at the boundary of senescence and programmed cell death in cucumber. *Plant Physiol.* **123**, 917–927
80. Liu, C. Y., Xu, H., and Graham, J. S. (1998) Cloning and characterization of an *Arabidopsis thaliana* cDNA (accession no. AF062640) homologous to the matrix metalloproteinases (PGR98–130). *Plant Physiol.* **117**, 1125–1127
81. Gamas, P., de Billy, F., and Truchet, G. (1998) Symbiosis-specific expression of two *Medicago truncatula* nodulin genes, MtN1 and MtN13, encoding products homologous to plant defense proteins. *Mol. Plant-Microbe Interact.* **11**, 393–403
82. Aziz, N., Paiva, N. L., May, G. D., and Dixon, R. A. (2005) Profiling the transcriptome of alfalfa glandular trichomes. *Planta* **221**, 28–38
83. Turner, J. G., Ellis, C., and Devoto, A. (2002) The jasmonate signal pathway. *Plant Cell* **14**, S153–S164
84. Schwechheimer, C., Serino, G., Callis, J., Crosby, W. L., Lyapina, S., Deshaies, R. J., Gray, W. M., Estelle, M., and Deng, X.-W. (2001) Interactions of the COP9 signalosome with the E3 ubiquitin ligase SCFTIR1 in mediating auxin response. *Science* **292**, 1379–1382
85. Nagano, Y., Murai, N., Matsuno, R., and Sasaki, Y. (1993) Isolation and characterization of cDNAs that encode eleven small GTP-binding proteins from *Pisum sativum*. *Plant Cell Physiol.* **34**, 447–455
86. Yamagata, H., Harada, F., Bowler, C., Teruo, I., and Aizono, Y. (1997) Molecular cloning of two cDNAs encoding small GTP-binding proteins from dark-adapted soybean suspension-cultured cells (accession nos. U58853, U58854) (PGR97–181). *Plant Physiol.* **115**, 1730
87. Debeaujon, I., Peeters, A. J. M., Leon-Kloosterziel, K. M., and Koornneef, M. (2001) The TRANSPARENT TESTA12 gene of *Arabidopsis* encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *Plant Cell* **13**, 853–872
88. Feng, S., Ma, L., Wang, X., Xie, D., Dinesh-Kumar, S. P., Wei, N., and Deng, X. W. (2003) The COP9 signalosome interacts physically with SCFCO1 and modulates jasmonate responses. *Plant Cell* **15**, 1083–1094
89. Lindsay, W. P., McAlister, F. M., Zhu, Q., He, X. Z., Droge-Laser, W., Hedrick, S., Doerner, P., Lamb, C., and Dixon, R. A. (2002) KAP-2, a protein that binds to the H-box in a bean chalcone synthase promoter, is a novel plant transcription factor with sequence identity to the large subunit of human Ku autoantigen. *Plant Mol. Biol.* **49**, 503–514
90. Hopfner, K. P., Putnam, C. D., and Tainer, J. A. (2002) DNA double-strand break repair from head to tail. *Curr. Opin. Struct. Biol.* **12**, 115–122
91. Gunawardena, U., and Hawes, M. C. (2002) Tissue specific localization of root infection by fungal pathogens: role of root border cells. *Mol. Plant-Microbe Interact.* **15**, 1128–1136
92. Miyasaka, S. C., and Hawes, M. C. (2001) Possible role of root border cells in detection and avoidance of aluminum toxicity. *Plant Physiol.* **125**, 1978–1987
93. Seregelyes, C., Mustardy, L., Ayaydin, F., Sass, L., Kovacs, L., Endre, G., Lukacs, N., Kovacs, I., Vass, I., Kiss, G. B., Horvath, G. V., and Dudits, D. (2000) Nuclear localization of a hypoxia-inducible novel non-symbiotic hemoglobin in cultured alfalfa cells. *FEBS Lett.* **482**, 125–130
94. Cao, Y., and Schubert, K. R. (2001) Molecular cloning and characterization of a cDNA encoding soybean nodule IMP dehydrogenase. *Biochim. Biophys. Acta* **1520**, 242–246
95. Shi, H., Xiong, L., Stevenson, B., Lu, T., and Zhu, J. K. (2002) The *Arabidopsis* salt overly sensitive 4 mutants uncover a critical role for vitamin B6 in plant salt tolerance. *Plant Cell* **14**, 575–588
96. Shi, H., and Zhu, J. K. (2002) SOS4, a pyridoxal kinase gene, is required for root hair development in *Arabidopsis*. *Plant Physiol.* **129**, 585–593
97. Achnine, L., Huhman, D., Farag, M., Sumner, L., Blount, J., and Dixon, R.

- (2005) Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J.* **41**, 875–887
98. Watson, B., Lei, Z., Dixon, R., and Sumner, L. (2004) Proteomics of *Medicago sativa* cell walls. *Phytochemistry* **65**, 1709–1720
99. Liu, C., Huhman, D., Sumner, L., and Dixon, R. (2003) Regiospecific hydroxylation of isoflavones by cytochrome p450 81E enzymes from *Medicago truncatula*. *Plant J.* **36**, 471–484
100. Chen, F., Duran, A., Blount, J., Sumner, L., and Dixon, R. (2003) Profiling phenolic metabolites in transgenic alfalfa modified in lignin biosynthesis. *Phytochemistry* **64**, 1013–1021
101. Suzuki, H., Achnine, L., Xu, R., Matsuda, S. P., and Dixon, R. A. (2002) A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. *Plant J.* **32**, 1033–1048