Proteomic Analysis of Common Bean (*Phaseolus vulgaris* L.) by Two-Dimensional Gel Electrophoresis and Mass Spectrometry

Savithiry S. Natarajan[a,*], M.A. Pastor-Corrales[a], Farooq H. Khan[a] and Wesley M. Garrett[b]

[a]*USDA-ARS, Soybean Genomics and Improvement Laboratory, Beltsville, MD 20705, USA*

[b]*USDA-ARS, Animal Bioscience and Biotechnology Laboratory, Beltsville, MD 20705, USA*

Abstract: The modern cultivated common bean (*Phaseolus vulgaris*) has evolved from wild common beans distributed in Central America, Mexico and the Andean region of South America. It has been reported that wild common bean accessions have higher levels of protein content than the domesticated dry bean cultivars. However, there is limited proteomic analysis of wild or domesticated common beans. In this investigation, we reported a proteomic analysis of a wild bean from Mexico. We utilized the TCA/Acetone method for protein extraction, two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) for protein separation, and subsequent mass spectrometry (MS) for protein identification. Proteins were separated in the first dimension using pH range from 4-7. A total of 237 protein spots from the second dimension were isolated, digested with trypsin, and analyzed by MALDI/TOF/TOF mass spectrometry. We identified 141 protein spots by searching NCBI non redundant databases using the Mascot search engine and found a total of 43 unique proteins. Gene Ontology (GO) analysis was employed to understand the molecular processes in which the identified common bean proteins are involved. The majority of proteins are involved in binding (41.5%) and catalytic activity (35.8%), followed by nutrient reservoir activity (7.5%), antioxidant activity (1.9%), transporter activity (3.8%), enzyme regulator activity (3.8%), structural molecule activity (1.9%), and electron carrier activity (3.8%). The results indicate that TCA/Acetone extraction and 2D-PAGE is efficient in separating common bean proteins. Further identification of these proteins by MS demonstrates the quality of this protein extraction method. These results will be useful for the development of value added common beans by alteration of protein components.

Keywords: Common bean, seed protein, proteomics, two-dimensional gel electrophoresis, MALDI-TOF/TOF, mass spectrometry.

INTRODUCTION

Dry seed or the green pods of common beans (*Phaseolus vulgaris* L.) are by far the most important legume for direct human consumption in the world. The common bean contributes significantly to the diets of millions of people in countries of Eastern, Central and Southern Africa, Latin America, Asia, and the Middle East. The protein in dry seeds of common bean is rich in lysine but low in the sulfur containing amino acids, namely cysteine and methionine. The common bean is also a significant source of complex carbohydrates, fiber, and minerals. The color, size, and shape of seeds of dry beans vary considerably and determine the so called bean market classes. At least 10 dry bean market classes are produced in the United States. These include the pinto, black, great northern, navy, dark and light red kidney varieties.

Dry beans and other legumes accumulate large amounts of protein during seed development. Typically, most dry beans contain 15 to 25% protein on a dry weight basis [1]. Proteins in common bean have been assigned to two groups, the salt-soluble globulins and the water-soluble albumins that account, respectively, for up to 45 to 70% and 10 to 30% of the total protein [1]. Dry bean albumins are typically composed of several different proteins, including lectins and enzyme inhibitors that are considered general defensive compounds.

Wild common beans, distributed from the highlands of South America to northern Mexico, are the progenitors of the cultivated or domesticated common bean landraces, which in turn gave origin to the currently grown common bean cultivars. The common bean is comprised of two distinct gene pools from Mesoamerica and the Andes, which predates domestication [2]. Some wild beans have been reported as having higher protein content than cultivated varieties, as well as similar protein digestibility and better levels of some essential amino acids.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is a useful proteomic tool, in which proteins are separated based on their isoelectric point (pI) in the first dimension, and subsequently based on their molecular weight (MW) in the second dimension, have been widely used to study cellular function in prokaryotic and eukaryotic organisms. A number of proteomic studies on different organs and organelles of soybean and other crops have been published [3-6]. Recent advancements in

*Address correspondence to this author at the Soybean Genomics and Improvement Laboratory, PSI-ARS-USDA, 10300 Baltimore Avenue, Beltsville, MD 20705, USA; Tel: 301-504-5258; Fax: 301-504-5728;
E-mail: savi.natarajan@ars.usda.gov

ISSN: 1814-8085 / E-ISSN: 1927-5129/13

separation methods, instrumentation, mass accuracy and sensitivity have led to the universal adoption of this method.

Variation of different classes of proteins in different genotypes of soybean seed, including anti-nutritional, allergen and storage proteins, has been investigated using 2D-PAGE in our laboratory [5]. In the present study, we describe the extraction, separation, and identification of different classes of common bean proteins. The Mascot search engine was used to search NCBI databases for protein identification. The identified proteins were clustered into cohesive groups based on their biochemical functions. Our findings are useful for understanding the various physiological functions of common bean proteins and useful for improvement of common beans protein quality through breeding programs as well as genetic manipulations.

METHODS

Protein Extraction from Common Bean (*Phaseolus Vulgaris*) Seeds

A modified TCA/Acetone method [4] was utilized to extract protein from common bean seeds. Seeds were first pulverized to powder in liquid nitrogen using a mortar and pestle. 150 µg of seed powder was homogenized in 4.5mL precipitation solution containing 10% (w/v) TCA in acetone with 0.07% (v/v) 2-mercaptoethanol. The protein was allowed to precipitate for 24 hrs at -20°C. To minimize the inclusion of lipids and pigments, the precipitated protein was then washed 4 times with acetone containing 0.07% (v/v) 2-mercaptoethanol, with each wash followed by centrifugation at 20800g for 20 minutes at 4°C. The supernatant from each wash was discarded. Subsequently, the precipitate was dried through the use of a speed vacuum for 30 minutes, and then resolubilized in 1.5mL lysis buffer [7M urea, 2M thiourea, 4% (w/v) CHAPS, 0.5% (w/v) DTT]. The mixture was homogenized using a sonicator for 10 minutes, and subsequently centrifuged at 20800g for 20 minutes at 4°C. Protein concentration of the supernatant, which contained the solubilized protein, was determined as describe earlier [4]. An appropriate aliquot consisting of 200ug protein was used for 2D-PAGE assay.

2D-PAGE Analysis

An IPGphor apparatus (GE Healthcare, Piscataway, NJ) was used for IEF with immobilized pH gradient (IPG) strips (pH 4.0 - 7.0, linear gradient, 13 cm). The IPG strips were rehydrated 12 hrs with 250µL rehydration buffer (7M urea, 2Mthiourea, 4% CHAPS, 2% pharmalyte, 0.002% bromophenol blue) containing 200 µg protein. The voltage settings for IEF was 500 V for 1 hr, 1000 V for 1 hr, 5000 V for 1 hr, and 8000 V to a total 46.86 kVh. Following electrophoresis, the protein in the strips was denatured with equilibration buffer (50mM Tris-HCl pH 8.8, 6M urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue, 1% DTT) and then incubated with the same buffer containing 2.5% iodoacetamide instead of DTT for 30 min at room temperature. The second dimension electrophoresis was performed on a 12.5% polyacrylamide gel using a Hoefer SE 600 Ruby electrophoresis unit (GE Healthcare, Piscataway, NJ). The gels were then stained with Colloidal Coomassie G-250 for the duration of 24 hours. After de-staining with ddH20 for several hours, gels were scanned using a GE ImageScanner III (GE Healthcare, Piscataway, NJ). Common bean proteins resolved by 2D PAGE were excised with a 1.5 mm Spot Picker (The Gel Company, San Francisco, CA, USA).

In-Gel Digestion of Protein Spots

Protein digestion was performed as described previously [4]. Spots were excised from the stained gel and washed with 50% acetonitrile containing 25 mM ammonium bicarbonate to remove the dye. The gel plug was dehydrated with 100% acetonitrile, dried under vacuum and incubated overnight at 37°C with 20 µL of 10 µg/mL porcine trypsin (Promega, sequencing grade, Madison, WI) in 20 mM ammonium bicarbonate for protein digestion. The resulting tryptic fragments were eluted by diffusion into 50% acetonitrile and 5% trifluoroacetic acid. A shaker was used to facilitate the diffusion. The extract was vacuum dried and the pellet was dissolved in 50% acetonitrile and 0.1% trifluoroacetic acid.

Mass Spectrometry

The mass spectrometer used was an AB SCIEX TOF/TOF™ 5800 System (AB SCIEX, Framingham, MA) operated in positive ion reflector mode to analyze tryptic peptides. Prior to analysis of unknowns, a plate model calibration was run to optimize mass accuracy and to update the instrument's default calibration parameters. Thirty fmol of a commercially prepared tryptic digest of bovine serum albumin (Michrom Bioresources, Inc. Auburn, CA) was spotted onto the 13 calibration wells of the sample plate and 5 peptides

with masses in the range of 927.493 m/z to 1881.905 m/z were used for the calibration. Samples were spotted on the plate, co-crystallized with a 5 mg/ml concentration of α-cyanohydroxycinnamic acid (CHCA) matrix prepared in 70% acetonitrile containing 0.1% trifluoroacetic acid.

The instrument was operated in batch mode during peptide analysis, which entails first performing an MS survey scan on all spots of interest, followed by sequential MS/MS analysis of peaks detected in the MS scan. Acquisition of MS/MS data was controlled by an interpretation method that acquired MS/MS spectra on the strongest precursors first on up to 100 precursors detected in the MS scan. An exclusion mass list was prepared to prevent MS/MS analysis of common human keratin contaminant and minor porcine trypsin autolysis peaks. MS spectra for both standards and unknowns were acquired in positive ion reflector mode with 400 shots of a 349 nm Nd: YAG laser operating at 404 Hz. MS/MS spectra were also acquired in positive ion reflector mode with 250-1000 laser shots firing at a rate of 1010 Hz. Collision energy was set to 1kV and collision induced dissociation (CID) was enabled with air as the collision gas in the CID cell. When possible, known trypsin autolysis peaks at m/z 842.51 and 2,211.10 were used to internally calibrate the MS spectra. Mascot Distiller version 2.3.2.0 (matrix science) was used to create searchable peak lists from the individual exported MS/MS spectral files. For TOF/TOF data, a custom Perl script provided by Matrix Science technical support was used to concatenate multiple individual ms/ms peak list files into a single sample specific searchable peak list.

Protein identification was performed using the Mascot search engine (http://www.matrixscience.com), which uses a probability based scoring system. NCBI non-redundant databases were selected as the primary databases to be searched. The following parameters for database searches were used: masses were set to monoisotopic; trypsin was set as the digestive enzyme allowing 1 missed cleavage; peptide tolerance of ± 50 parts per million (ppm); MS/MS tolerance of ±0.6 Da; inclusion of peptide charges of 1+; fixed modification of carbamidomethylation of cysteine; variable modifications included oxidation of methionine and N-terminal pyroglutamic acid from glutamic acid or glutamine. The samples were identified using the taxonomy filter of Viridiplantae (green plants). A minimum MASCOT score (combined ion score) of 60 with at least one significantly matched peptide was required for a given result to qualify as a positive identification. Samples not identified through NCBI were subjected to a BLAST sequence similarity search on the UNIProt protein sequence database, where a maximum E-value of 1.00E-100 along with a minimum sequence similarity of 65% were used as cut-offs for a positive protein identification. Subsequently, unidentifiable samples with abundant MS/MS spectra were searched against a downloaded FASTA formatted database from Phytozome (Pvulgaris_186_peptide. fasta, 30,721 sequences as of 1/10/2012), downloaded from http://www.phytozome.net/ and uploaded on a local Mascot (Matrix Science, London, UK; version 2.3.02) server for searching of protein spots. Regarding the Phytozome peptide database, these sequence data were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/ in collaboration with the user community. Positive identification of proteins by MS/MS analysis required a minimum of two unique peptides, with at least one peptide having a significant ion score.

RESULTS AND DISCUSSION

Separation of Common Bean Proteins

We employed a modified TCA/acetone method to extract common bean proteins according to Natarajan *et al*. [4]. Three biological replicates of protein extracts were separated by 2D-PAGE. Initial analyses were performed with immobilized pH gradient (IPG) strips with a pH range of 3 to 10 (data not shown). Upon observation that the region from pH 4 to 7 was a highly dense area we, thereafter, used pH 4 to7 IPG strips to improve the resolution of protein spots and further their identification (Figure **1**). A representative 2D-PAGE protein pattern of common bean protein is presented in Figure **1**. Two hundred and thirty seven spots were excised from the 2D-PAGE gels and digested with trypsin. The tryptic digests were purified and then analyzed by MALDI/TOF/TOF mass spectrometry. The dynamic range of several proteins including storage proteins is abundant, similar to soybean storage proteins and can mask clear detection of other low abundant proteins. However, we were able to identify a significant number of low abundant proteins in our common bean seed samples.

Identification of Common Bean Proteins

Common bean protein information such as assigned protein spot number, theoretical isoelectric point (pI), and molecular weight (Mr), protein identity and its original species, number of peptides matched,

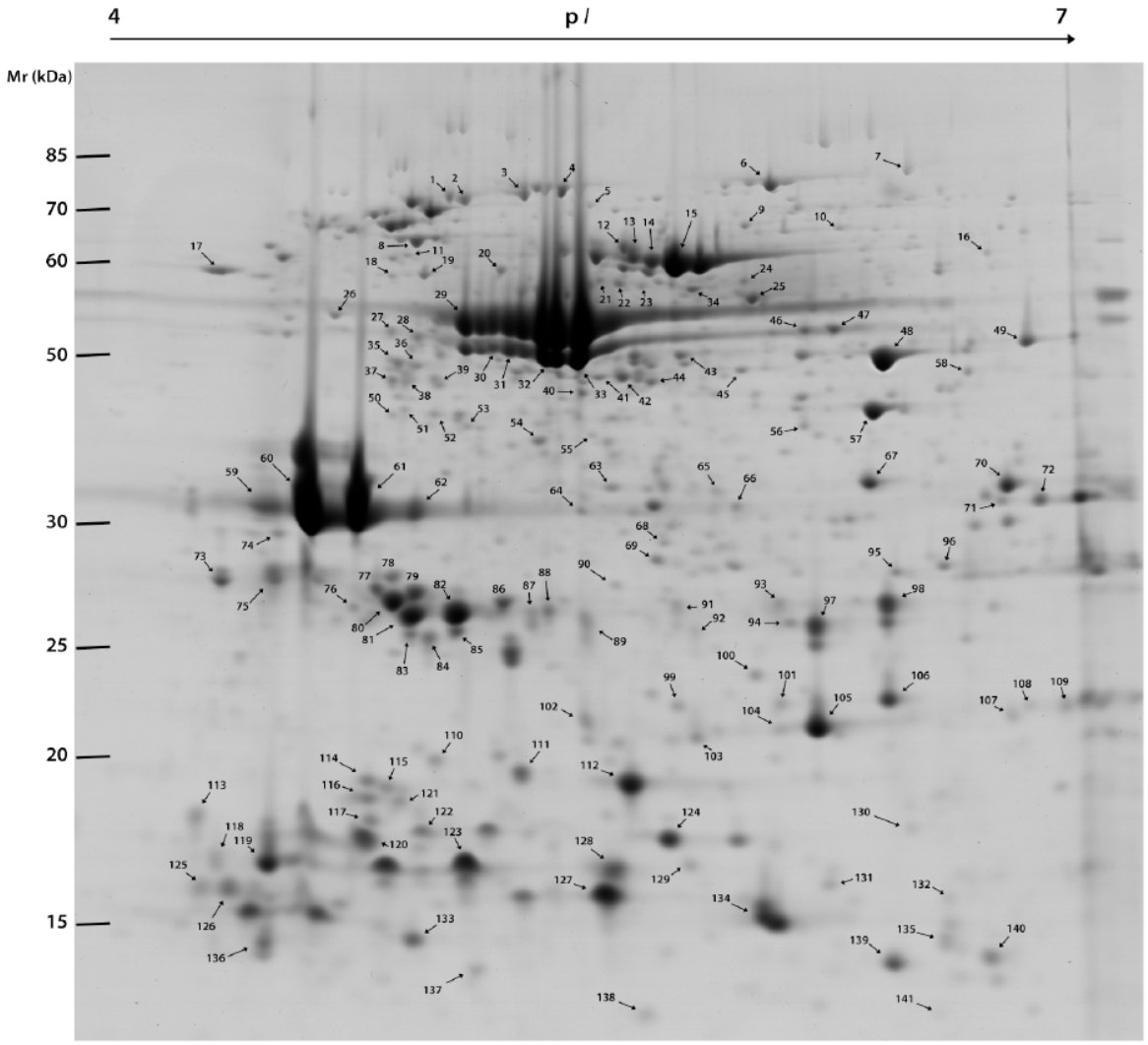

Figure 1: 2-DE map of *Phaseolus vulgaris* seed protein. Proteins (200µg) were separated in a gradient of pH4-7 for the first dimension, then resolved using 12.5% SDS-PAGE in the second dimension. The gels were stained with Colloidal Coomassie Blue G-250. A total of 141 protein spots were excised and identified with MALDI TOF-TOF MS/MS. Detailed information about the protein properties and identification is listed in Table **1**.

percentage sequence coverage, MOWSE score, database searched, and accession number of the best match for the 141 identified protein spots are listed in Table **1**. Of the 237 protein spots processed, 141 proteins were successfully identified by querying the NCBI nr database using the Mascot search engine. Some of the identified proteins that had a significant score matching a hypothetical or predicted or unknown listing in NCBI were subsequently subjected to a homology search against the UniProt (www.uniprot.org) database using BLAST (Basic Local Alignment Search Tool).

Storage Proteins

Previous reports are available on the storage proteins of soybean, buckwheat grains, tomato and other crop seeds [5]. To our knowledge, the existing reference map of common bean proteins is limited. Common bean seeds contain approximately 15 - 25% protein on a dry weight basis. Sathe [1] reported that a large portion of bean protein is made up of globulins, which are salt soluble and account for up to 10-30%, and water soluble albumins that account for up to 10-30% of the total proteins. *Phaseolus vulgaris* crude protein contains 12-30% of albumin, 54-79% of globulin, 2-4% of prolamin and 20-30% of glutelin. Reported albumins are typically composed of several different proteins including lectins and enzyme inhibitors [1]. Louis *et al*. [7] reported that albumin-1 is associated with an insecticidal activity in seed extracts of common bean and reported the presence of multiple isoforms using proteomics. Seed albumin-1 is recognized to be highly polymorphic in common bean

Table 1: Proteins Identified from Common Bean (*Phaseolus vulgaris*) Seeds by MALDI TOF-TOF MS/MS

ID	Protein Identificaiton [species]	Calc pI/Mr	MS Score	# PM	% SC	gi / Uniprot Accession No.	Blast E-Value	Blast % Identity
1	Endoplasmic reticulum HSC70-cognate binding protein precursor [*Glycine max*]	5.15 / 73822	423	6	14	2642238		
2	luminal binding protein (BiP) [*Arabidopsis thaliana*]	5.08 / 73732	230	2	9	1303695		
3	Heat-shock protein [*Medicago truncatula*]	5.28 / 71894	237	2	11	357480003		
4	Phaseolin, alpha-type [*Phaseolus vulgaris*]	5.25 / 49241	104	2	9	130169		
5	Phaseolin, alpha-type [*Phaseolus vulgaris*]	5.25 / 49241	63	1	4	130169		
6	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	453	4	15	312982406		
7	Cobalamine-independent methionine synthase [*Solenostemon scutellarioides*]	6.17 / 87059	167	2	7	974782		
8*	Protein disulfide isomerase [*Glycine max*]	5.35 / 62467	274	3	9	B1Q2X4	0	91
9	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	190	2	9	312982406		
10	Lipoxygenase [*Phaseolus vulgaris*]	5.94 / 97706	81	1	4	11991840		
11	Group 3 late embryogenesis abundant protein [*Phaseolus vulgaris*]	5.80 / 50724	335	5	12	75708857		
	Phaseolin [*Phaseolus vulgaris*]	5.35 / 48448	92	1	8	403596		
12	Group 3 late embryogenesis abundant protein [*Phaseolus vulgaris*]	5.80 / 50724	118	1	9	75708857		
13	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	411	4	12	312982406		
	Group 3 late embryogenesis abundant protein [*Phaseolus vulgaris*]	5.80 / 50724	185	1	9	75708857		
14	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	509	7	15	312982406		
15	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	847	10	22	312982406		
	Group 3 late embryogenesis abundant protein [*Phaseolus vulgaris*]	5.80 / 50724	197	2	11	75708857		
16	F1-ATPase alpha subunit [*Hydrothrix gardneri*]	7.08 / 41915	131	2	12	34539327		
17	Calreticulin [*Medicago truncatula*]	4.41 / 48595	80	1	2	357441869		
18	Beta-tubulin [*Glycine max*]	5.63 / 46348	299	2	17	1351202		
19	UTP--glucose-1-phosphate uridylyltransferase 1 [*Arabidopsis thaliana*]	5.73 / 52058	168	2	4	15237947		
20	ATP synthase subunit beta [*Medicago truncatula*]	5.86 / 121227	703	8	12	357444609		
	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	114	2	9	403594		
21	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	383	5	18	403594		
	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	114	1	4	312982406		
22	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	281	3	8	312982406		
	Phaseolin [*Phaseolus vulgaris*]	5.42 / 48533	181	1	13	403594		
23	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	153	1	8	312982406		
	Phaseolin, alpha-type [*Phaseolus vulgaris*]	5.25 / 49241	148	1	13	130169		
24	Phaseolin, alpha-type [*Phaseolus vulgaris*]	5.25 / 49241	214	3	13	130169		
25	Formate dehydrogenase [*Phaseolus vulgaris*]	6.47 / 41524	769	6	36	270342112		
	Phaseolin, alpha-type [*Phaseolus vulgaris*]	5.25 / 49241	230	4	11	130169		
26	Heat-shock protein [*Medicago truncatula*]	5.28 / 71894	107	1	3	357480003		
27	Alpha-phaseolin [*Phaseolus vulgaris*]	5.42 / 48533	578	5	25	295832		
28	Alpha-phaseolin [*Phaseolus vulgaris*]	5.42 / 48533	556	6	26	295832		
29	Alpha-phaseolin [*Phaseolus vulgaris*]	5.25 / 49241	1243	15	41	130169		
30	Alpha-phaseolin [*Phaseolus vulgaris*]	5.25 / 49241	1020	14	36	130169		
31	Alpha-phaseolin [*Phaseolus vulgaris*]	5.25 / 49241	1114	12	42	130169		

(Table 1). Continued.

ID	Protein Identificaiton [species]	Calc pI/Mr	MS Score	# PM	% SC	gi / Uniprot Accession No.	Blast E-Value	Blast % Identity
32	Alpha-phaseolin [*Phaseolus vulgaris*]	5.25 / 49241	1320	17	41	130169		
33	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	1355	16	42	403594		
34	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	280	3	10	312982406		
	Phaseolin [*Phaseolus vulgaris*]	5.35 / 48448	151	3	10	403596		
35	Phaseolin [*Phaseolus vulgaris*]	5.35 / 48448	661	7	29	403596		
36	Alpha-phaseolin [*Phaseolus vulgaris*]	5.42 / 48533	565	7	33	295832		
37	Phaseolin [*Phaseolus vulgaris*]	5.35 / 48448	639	6	24	403596		
38	Alpha-phaseolin [*Phaseolus vulgaris*]	5.42 / 48533	605	7	29	295832		
39	Alpha-phaseolin [*Phaseolus vulgaris*]	5.42 / 48533	632	8	27	295832		
40	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	142	1	6	312982406		
41	Probable protein disulfide-isomerase A6 [*Medicago sativa*]	5.44 / 40809	71	1	3	729442		
42	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	393	7	11	312982406		
43	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	284	3	9	312982406		
44	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	132	2	4	312982406		
45	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	137	2	6	312982406		
46	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	278	2	13	403594		
	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	125	1	6	312982406		
47	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	261	2	9	312982406		
	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	148	1	9	403594		
48*	Lea Protein [*Glycine max*]	6.10 / 41003	482	8	18	Q39873	1.00E-154	65
49	Alcohol dehydrogenase 2b [*Gossypium hirsutum*]	6.23 / 41748	209	3	8	1263291		
50	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	147	2	9	403594		
51	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	246	3	13	403594		
52	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	298	3	14	403594		
53*	Fructokinase [*Medicago truncatula*]	5.07 / 35602	258	4	20	G7IHV2	1.00E-169	92
54	Mannose lectin FRIL [*Phaseolus vulgaris*]	5.38 / 31083	437	5	28	6822274		
55	60S acidic ribosomal protein P0 [*Glycine max*]	5.15 / 34194	63	1	4	351725493		
56	Ribulose bisphosphate carboxylase large chain [*Adoxa moschatellina*]	6.13 / 52159	214	3	7	131899		
57	IAA-protein conjugate [*Phaseolus vulgaris*]	6.20 / 35494	582	8	36	11363146		
58	Glutamine synthetase PR-1 [*Phaseolus vulgaris*]	5.78 / 39265	181	3	8	121334		
59	Phytohemagglutinin [*Phaseolus vulgaris*]	4.95 / 29806	595	7	28	19744132		
60	Phytohemagglutinin [*Phaseolus vulgaris*]	4.95 / 29806	898	9	44	19744132		
61	Phytohemagglutinin [*Phaseolus vulgaris*]	5.34 / 29462	727	8	36	501100		
62	Phytohemagglutinin [*Phaseolus vulgaris*]	5.34 / 29462	556	7	29	501100		
63	Phaseolin, beta type [*Phaseolus vulgaris*]	5.29 / 47536	99	2	6	130170		
64	Formate dehydrogenase [*Phaseolus vulgaris*]	6.47 / 41524	238	2	10	270342112		
65	Phaseolin [*Phaseolus vulgaris*]	5.35 / 48448	154	2	6	403596		
66	Alpha-phaseolin [*Phaseolus vulgaris*]	5.42 / 48533	198	3	9	295832		
67*	Yieldin [*Vigna unguiculta*]	5.87 / 36870	126	2	6	Q9MBC9	1.00E-136	74
68	Alpha-phaseolin [*Phaseolus vulgaris*]	5.42 / 48533	125	1	9	295832		

(Table 1). Continued.

ID	Protein Identificaiton [species]	Calc pI/Mr	MS Score	# PM	% SC	gi / Uniprot Accession No.	Blast E-Value	Blast % Identity
69	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	188	2	6	312982406		
70*	Yieldin [*Vigna unguiculta*]	5.40 / 35566	118	2	6	Q9MBC9	1.00E-133	74
71	Dehydrin [*Vigna unguiculata*]	5.97 / 26508	65	1	5	6358640		
72	Dehydrin [*Vigna unguiculata*]	5.97 / 26508	118	2	5	6358640		
73	Phytohemagglutinin [*Phaseolus vulgaris*]	4.95 / 29806	391	5	24	19744132		
74	Lectin [*Phaseolus vulgaris*]	4.83 / 29608	202	2	17	19773406		
	Phytohemagglutinin [*Phaseolus coccineus*]	5.80 / 29504	153	2	10	19577338		
75	Alpha-amylase inhibitor like protein [*Phaseolus vulgaris*]	4.95 / 28882	117	2	7	6456428		
76	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	412	4	15	403594		
77	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	231	2	10	403594		
	Albumin-2 [*Phaseolus vulgaris*]	6.90 / 25604	86	1	5	312982408		
78	Phytohemagglutinin [*Phaseolus vulgaris*]	5.34 / 29462	290	5	17	501100		
79	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	691	7	19	403594		
80	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	798	8	20	130169		
81	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	809	9	20	403594		
82	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	956	10	27	403594		
83	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	436	4	18	403594		
84	Phytohemagglutinin [*Phaseolus vulgaris*]	4.95 / 29806	436	5	20	19744132		
85	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	777	8	27	403594		
86	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	199	2	12	403594		
87	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	122	1	4	403594		
88	Phytohemagglutinin [*Phaseolus vulgaris*]	4.95 / 29806	103	1	6	19744132		
89	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	355	5	16	130169		
90	Superoxide dismutase [Fe], chloroplastic precursor [*Glycine max*]	5.60 / 27881	93	2	9	351721352		
91	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	90	1	4	403594		
92	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	226	3	9	130169		
93	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	313	3	10	130169		
94	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	193	3	6	312982406		
95	Triose-phosphate isomerase [*Phaseolus vulgaris var. nanus*]	5.87 / 27411	423	4	24	57283985		
96	Alcohol dehydrogenase 2b [*Gossypium hirsutum*]	6.23 / 41748	143	2	6	1263291		
97	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	825	11	28	403594		
98	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	757	10	25	130169		
	Glutathione transferase [*Phaseolus vulgaris*]	5.84 / 24821	131	1	13	371572988		
99	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	402	4	14	130169		
100	Albumin-2 [*Phaseolus vulgaris*]	6.90 / 25604	243	3	18	312982408		
101	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	349	5	10	130169		
102	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	548	5	20	130169		
103	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	569	8	22	403594		
104	Phaseolin, beta type [*Phaseolus vulgaris*]	5.29 / 47536	115	1	6	130170		
105	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	998	13	32	403594		
106	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	967	12	30	130169		

(Table 1). Continued.

ID	Protein Identificaiton [species]	Calc pI/Mr	MS Score	# PM	% SC	gi / Uniprot Accession No.	Blast E-Value	Blast % Identity
107	Formate dehydrogenase [*Phaseolus vulgaris*]	6.47 / 41524	274	3	15	270342112		
108	Formate dehydrogenase [*Phaseolus vulgaris*]	6.47 / 41524	151	1	13	270342112		
109	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	304	3	5	312982406		
110	Mannose lectin FRIL [*Phaseolus vulgaris*]	5.38 / 31083	332	3	15	6822274		
111	Albumin-2 [*Phaseolus vulgaris*]	6.90 / 25604	346	4	25	312982408		
112	Mannose lectin FRIL [*Phaseolus vulgaris*]	5.38 / 31083	396	3	15	6822274		
113	Phytohemagglutinin [*Phaseolus vulgaris*]	4.95 / 29806	128	1	6	19744132		
114	Phytohemagglutinin [*Phaseolus vulgaris*]	5.34 / 29462	73	1	3	501100		
115	Mannose lectin FRIL [*Phaseolus vulgaris*]	5.38 / 31083	273	3	15	6822274		
116	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	173	1	6	130169		
117	Phytohemagglutinin [*Phaseolus vulgaris*]	5.34 / 29462	231	3	18	501100		
118	Alpha-amylase inhibitor, PHA-I beta subunit [*Phaseolus vulgaris*]	4.70 / 15395	297	4	49	1911780		
119	Alpha-amylase inhibitor, PHA-I beta subunit [*Phaseolus vulgaris*]	4.70 / 15395	281	3	24	1911780		
120	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	457	4	19	130169		
121	Phaseolin [*Phaseolus vulgaris*]	5.42 / 47525	547	5	19	403594		
122	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	147	1	9	130169		
123	Double-headed trypsin inhibitor [*Phaseolus vulgaris*]	6.38 / 12276	121	2	25	21304454		
124	Low molecular weight heat-shock protein [*Papaver somniferum*]	6.36 / 23917	116	1	9	507209		
125	Phytohemagglutinin [*Phaseolus vulgaris*]	5.34 / 29462	162	1	14	501100		
126	Phytohemagglutinin [*Phaseolus vulgaris*]	5.34 / 29462	463	5	27	501100		
127	Mannose lectin FRIL [*Phaseolus vulgaris*]	5.38 / 31083	130	2	11	6822274		
	Double-headed trypsin inhibitor [*Phaseolus vulgaris*]	6.38 / 12276	81	1	16	21304454		
128	Mannose lectin FRIL [*Phaseolus vulgaris*]	5.38 / 31083	153	2	15	6822274		
129	17.3 kDa class II heat shock protein [*Solanum pervianum*]	6.32 / 17311	192	2	20	75279028		
130	Granule-bound starch synthase I [*Phaseolus vulgaris*]	6.43 / 67607	290	3	10	5441242		
131	Late embryogenesis abundant protein [*Gossypium hirsutum*]	5.55 / 11965	68	1	10	18499		
132	Mannose lectin FRIL [*Phaseolus vulgaris*]	5.38 / 31083	80	1	6	6822274		
133	Phytohemagglutinin [*Phaseolus vulgaris*]	4.95 / 29806	593	7	33	19744132		
134	Mannose lectin FRIL [*Phaseolus vulgaris*]	5.38 / 31083	374	4	18	6822274		
135	Mannose lectin FRIL [*Phaseolus vulgaris*]	5.38 / 31083	268	3	15	6822274		
136	Alpha-amylase inhibitor, PHA-I beta subunit [*Phaseolus vulgaris*]	4.70 / 15395	664	8	66	1911780		
137	Phaseolin, alpha type [*Phaseolus vulgaris*]	5.25 / 49241	209	2	10	130169		
138	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	301	3	4	312982406		
139	Em protein [*Glycine max*]	5.52 / 11485	114	1	24	351723669		
140	Legumin [*Phaseolus vulgaris*]	5.64 / 69081	452	5	8	312982406		
141	Nucleoside diphosphate kinase 1 [*Pisum savitum*]	5.94 / 16452	81	1	11	1346672		

*NCBI database search for these spots resulted in an 'unknown' or 'hypothetical' protein match. Therefore the resulting sequences were subjected to a sequence similarity search *via* BLAST under the UNIPROT database, in order to obtain proper identity of the protein. The Blast e-value and % identity are provided.
ID, spot number referred to in Figure **1**; calc pI/Mr, theoretical values for isoelectric point and molecular weight; MS Score, MOWSE Score; PM, the number of peptides matched; SC, sequence coverage by PMF using MALDI TOF-TOF MS/MS.

and other legumes and contains 8.4% of its residues as Cys and 4.2% as Met. Marsolais *et al*. [8] reported that mung bean albumin-2 is another sulfur-rich protein, contributing to a higher Cys content than what is seen in soybean. The mung bean seed albumin identified by mass spectrometry showed 1.8% of its residues as Cys

and 0.7% as Met. In common bean seed protein, we found three protein spots (# 77, 100, and 111) by 2D-PAGE that were identified as albumin-2. The second type of storage protein, the globulins, account for about 50-60% of the total protein in mature cotyledons of 5 homozygous common beans. As far back as 1898, Osborne and Campbell [9] separated the globulin protein in pea, *Pisum sativum*, into 2 major fractions called vicilin and legumin, according to their solubility in different concentrations of salt and by different heat coagulation properties. Millerd [10] reviewed the characterization of legumin and vicilin by analytical ultracentrifugation, gel electrophoresis, amino acid composition and carbohydrate content determination. Similar protein fractions could be extracted from other legume seeds including common bean and soybean. In soybean, legumin and vicilin together account for about 70% of the total protein content. Synthesis, sorting and storage of both legumin and vicilin in the developing seed have been extensively reviewed [11]. Both proteins are typically synthesized as preproproteins in the endoplasmic reticulum (ER), transported through the Golgi apparatus, and deposited in protein bodies (PBs), or into protein storage vacuoles (PSVs) that are later transformed into PBs.

Phaseolin, the main globulin fraction in the seed of French bean belongs to the vicilin or 7S fraction, and it represents 40-50% of the total seed nitrogen. However, the protein is deficient in methionine, cysteine and tryptophan and it is therefore a principle target for modification to improve bean protein quality. Phaseolin is modified with neutral sugars that contribute to the high level of variation in the molecular weight of its subunits. Phaseolin molecular diversity has been used as an indicator of evolution of the common bean domesticated in Central America and in the Andes. This diversity provides botanic, historical and archaeological information due to polymorphism, its environmental stability and biochemical complexity characteristics. Salmanowicz [12] used sodium dodecyl sulfate-capillary gel electrophoresis (SDA-CGE) and studied phaseolin storage protein variability in 44 wild and cultivated *Phaseolus vulgaris*. The author reported that 11 phaseolin profiles, revealing polypeptide subunit variation ranging from 45.6 kDa to 54.4 kDa. These polypeptides also differ in their isoelectric points.

We found multiple spots identified as phaseolin in our 2D-PAGE gels of common bean seed proteins (Figure **1**). Our results showed that 28 spots (# 11, 20, 22, 34, 47, 21, 33, 35, 37, 46, 50, 51, 52, 65, 76, 77, 79, 81, 82, 83, 85, 86, 87, 91, 97, 103, 105, 121) were identified as phaseolin , 29 protein spots (# 4, 5, 23, 24, 25, 27, 28, 29, 30, 31, 32, 36, 38, 39, 66, 68, 80, 89, 92, 93, 98, 99, 101, 102, 106, 116, 120, 122, 137) were identified as alpha-phaseolin, and 2 spots (# 63, 104) were identified as beta-phaseolin. As reported by Salmanowicz [12], our results also showed variation of molecular weight (Table **1**) ranging from 47-50 kDa, and isoelectric point (5.25 to 5.42). Recently De La Fuente *et al*. [13] used 2D-PAGE analysis and compared common bean proteins extracted by different methods. They selected 50 spots and identified them using mass spectrometry to demonstrate the quality of their 2D-analysis. They reported identification of 15 phaseolin protein spots of molecular weight range and isoelectric points in agreement with our identified phaseolin proteins. In another report [14], these authors characterized phaseolin protein diversity in cultivated common bean of different geographic origins such as Mesoamerican and Andean gene pools. They concluded that the majority of the phaseolin spot patterns are differentially glycosylated α and β-type polypeptides. Bollini and Vitale [15] reported that phaseolin is a family of proteins varying in MW, pI and polypeptide composition due to the proportion of each polypeptide present in the whole molecule.

Previously, we reported variation of 7S globulin subunits (beta-conglycinin) in soybean seed among various genotypes [5]. Montoya *et al*. [16] investigated the amino acid composition and the susceptibility to *in vitro* proteolysis of purified phaseolins in unheated and heated form. Subunit patterns of the purified phaseolins were investigated using SDS-PAGE and it was found that the subunits of phaseolin varied from 2-6 bands with specific MWs ranging from 54.7-41.1 kDa. The differences in MW and pI found in our study could reflect different gene sequences coding for two different polypeptide sub-families, such as alpha-phaseolin and beta- phaseolin, derived from the same ancestor as suggested by Slightom *et al*. [17] It could also be due to pre- and post- translational modifications that lead to the differentiation of phaseolin polypeptides.

Legumin is another storage protein of the globulin fraction and was first described in several species including *Phaseolus vulgaris*, *Pisum sativum*, *Vicia faba* [9]. Legumin, a hexameric 11S globulin (360 kDa), consists of the acidic alpha subunit of about 40 kDa and the basic beta subunit of about 20kDa. Both polypeptides were linked together by a disulfide bond. Derbyshire *et al*. [18] reported that legumin represents a major storage protein in soybean, broad bean and

pea seeds. The molecular heterogeneity of legumin is well documented and could be due to the flexible region of the 11S type proteins that are susceptible to proteolysis [19]. We found 21 protein spots (# 21, 46, 6, 9, 13, 14, 15, 22, 23, 34, 40, 42, 43, 44, 45, 47, 69, 94, 109, 138, 140) with a molecular weight of 69 kDa and an isoelectric point of 5.64 that were identified as legumin by our 2D-PAGE analysis,. Members of 11S legumins are relatively higher in sulfur amino acids, including cysteine, when compared with 7S globulins.

Defense Related Proteins

Many seeds contain defense proteins, including lectins and amylase inhibitors, which were first identified in the common bean, *P. vulgaris*. The genes for each of these proteins consist of a single locus in the *P. vulgaris* genome, and it is likely that these homologous genes have arisen by duplication of a single ancestral gene [20]. Lectins, which are carbohydrate-binding glycoproteins, are capable of specific recognition and reversible binding or cross-linking to carbohydrates, without altering their covalent structure. The lectins are also referred to as phytohemagglutinins (PHA), glycoprotein I [21], and have been identified in over 70 different legume seeds. Because lectins are often resistant to heat and proteolytic enzymes, including those of intestinal microflora, the effects of consumption of these proteins deserves special attention. The potential health benefits of lectins from legumes including antiproliferative, antitumor, immunomodulatory, antifungal, antiviral, anticancer and HIV-1 reverse transcriptase inhibitor activities have been extensively studied [22]. Lectins exhibit many interesting properties, such as specificity for human blood types, preferential agglutination of malignant cells, and mitogenic stimulation of lymphocytes [23]. Depending on the post-translational processing of the precursor, legume lectins can be single or double chain. Both types are synthesized in the endoplasmic reticulum and undergo co-translational removal of the signal peptide and in many cases also n-glycosylation. The phytohemagglutinins are abundant, accounting for about 10% of total protein of the cotyledons, and have been characterized in a number of cultivars of *P. vulgaris*. Among other possible functions, they are responsible for innate immunity and defense mechanisms, as well as interaction with symbionts including mycorrhizal fungi and nitrogen-fixing rhizobia [24]. Our 2D-PAGE showed fifteen protein spots identified as phytohemagglutinin (# 59, 60, 61, 62, 73, 74, 78, 84, 88, 113, 114, 117, 125, 126, 133), nine protein spots identified as mannose lectin FRIL (# 54, 110, 112, 115, 127, 128, 132, 134, and 135) and one spot identified as simply lectin (# 74). The MW and pI of lectin protein spots varied ranging from 29 - 31kDa and 4.83 to 5.80, respectively. Sathe [1] reported a 27-37kDa molecular weight variation of lectins in common bean, which is similar to our results.

Alpha-amylase inhibitor is a defense related protein that inhibits porcine pancreatic α-amylase activity, forming a 1:1 complex with the enzyme, and it is responsible for insect resistance, especially starch rich seeds. Mirkov *et al*. [20] reported that 2 different α-amylase 1 sequences were found in a wild accession of *P. vulgaris*. This bean is resistant to bean weevil and it is possible that this resistance results from specific α-amylase inhibitors. Amino acid sequencing of α-amylase purified from seed meal of *P.vulgaris* cv Rico 23 showed the presence of 2 different proteins, each with its own α and β subunits. In our common bean proteins 2D-PAGE study, we found four spots of α-amylase (# 75, 118, 119, and 136).

Another defense related protein family, double headed trypsin inhibitors, also called Bowman-Birk inhibitor (BBIs) consists of cysteine-rich protease inhibitors with molecular masses of about 8–16 kDa and high cysteine content. They were discovered first in soybean seeds and are the most widespread group in common bean. These protease inhibitors have two reactive sites on a single inhibitor molecule, and are involved in plant protection from pests and pathogens. They represent a form of storage of sulfur amino acids during dormancy. We found 2 spots (# 123, 127) of double- headed trypsin inhibitor with molecular weight of 12.3 kDa. BBIs identified in the *Fabaceae*, including soybean (*G. max*) and lima beans (*Phaseolus lunatus*), are encoded by a family of related genes. Many studies with *in-vitro* and *in-vivo* model systems and in human trials have provided evidence that these inhibitors have anti-inflammatory and anti-carcinogenic properties [25]. Lioi *et al*. [26] investigated the primary structure and anti-tryptic activity of the double headed BBIs protein family in cultivated and wild *Phaseolus* species.

Stress Related Proteins

Plants undergo a series of physiological, biochemical and molecular changes in response to adverse environmental conditions or stresses such as drought, low temperature or salinity. Following exposure to various stresses, plants produce a group of proteins that are presumed to play protective roles.

One group of such proteins belongs to the dehydrin super-family. These proteins accumulate during the latter stages of seed embryogenesis, and a subset of these proteins, termed Late Embryogenesis Abundant (LEAs) or Responsive to ABA (RABs), have some common features in their structure. Rosenberg *et al*. [27] reported these proteins simply as maturation proteins that slowly disappear during germination. Dehydrin family proteins are also found in vegetative organs, especially under stress conditions such as cold, drought, and high salinity, when they may be involved in binding or replacement of water under dehydration conditions [28]. In addition, these proteins help to maintain protein and membrane structure. Some LEA proteins can protect enzyme activities, or interact with sugars to prevent protein aggregation [29]. Overexpression of certain LEA genes in transgenic plants resulted in enhanced tolerance to abiotic stresses and tolerance to drought and high salinity [30]. Liu *et al*. [31] reported that LEA proteins such as GmPM1 and GmPM9 are metal binding proteins that may function in reducing oxidative damage induced by abiotic stress. Shih *et al*. [32] characterized two soybean LEA IV proteins, acidic (GmPM28) and basic (GmPM1), using circular dichroism and Fourier transform infrared spectrometry, and concluded that the LEA IV proteins are functional in the dry state and may stabilize desiccation-sensitive proteins and plasma membranes during dehydration.

In our study, we found six LEA proteins spots (# 11, 12, 13, 15, 48, and 131) and 2 spots (# 71, 72) of dehydrin distributed at various pHs (Table **1**). In our earlier publications, we reported the variation of LEA proteins and dehydrin among wild and cultivated soybean seeds [33]. It has been hypothesized that dehydrins function as surfactant molecules, acting synergistically with compatible solutes to prevent coagulation of colloids and a range of macromolecules. Aghaei *et al*. [34] reported the LEA proteins were up-regulated in response to salt stress in soybean.

Other stress related proteins, the heat shock proteins (HSPs), are synthesized in response to elevated temperature and during various developmental processes, including seed maturation. Hong and Vierling [35] and Wehmeyer *et al*. [36] described that HSP synthesis is developmentally regulated, being abundant in dry mature seeds and disappearing during germination. HSPs represent ~1% of the total protein in soybean seedlings. HSPs function as molecular chaperones, associated with protein folding, protein translocation and degradation, assembly of oligomeric proteins, modulation of receptor activities, mRNA protection, prevention of protein denaturation, and stress-induced aggregation, and post-stress ubiquitin and chaperonin-aided repair [37]. DeRocher and Vierling [38] reported that LEAs and HSPs are involved in protective functions throughout germination in pea seeds. We found 2 protein spots (# 3, 26) that were identified as heat shock protein, one spot (# 129) identified as a class II small heat shock protein and another spot (# 124) identified as a low molecular weight heat shock protein. Ahsan *et al*. [39] analyzed and compared HSP in soybean seedling, leaves, stem, and roots under heat stress using 2D analysis. They reported that differentially expressed HSPs and proteins involved in antioxidant defense were mostly up-regulated and proteins associated with photosynthesis, secondary metabolism, amino acid and protein synthesis were down- regulated in response to heat stress. Stupnikova *et al*. [40] demonstrated the function of pea mitochondrial HSP22 in protecting against cold or high temperature in maturing seed embryos.

Another protein falling under functional categories of stress and pathogen / disease-defense was identified as superoxide dismutase (# 90). Superoxide dismutase inhibits the auto-oxidation of epinephrine and the photochemical reduction of nitro-blue tetrazolium (NBT) in soybeans seeds. Its enzymatic activity is affected during the initial stages of seed germination. Scandalios *et al*. [41] reported that superoxide radicals increase dramatically in response to biotic and abiotic stresses, and also increase due to salinity.

Proteins Involved in Metabolism

Several identified protein spots from our study are involved in general metabolism. Most of these are involved in amino acid metabolism and we identified them as the following: glutamine synthetase (# 58) and methionine synthase (# 7) which is an integral membrane protein involved in signaling. Other identified proteins are involved in protein synthesis. These include 60S acidic ribosomal protein (# 55) and ATP synthase subunit beta (# 20).

Other proteins identified support primary metabolic processes and synthesis of nucleotides and other secondary compounds including: 2 spots identified as kinases (# 141, 53), 2 protein disulfide isomerase spots (# 8, 41) and a triose phosphate isomerase (# 95), 2 spots (# 49, 96) identified as alcohol dehydrogenase 2b, 4 spots (# 25, 64, 107, 108) of formate

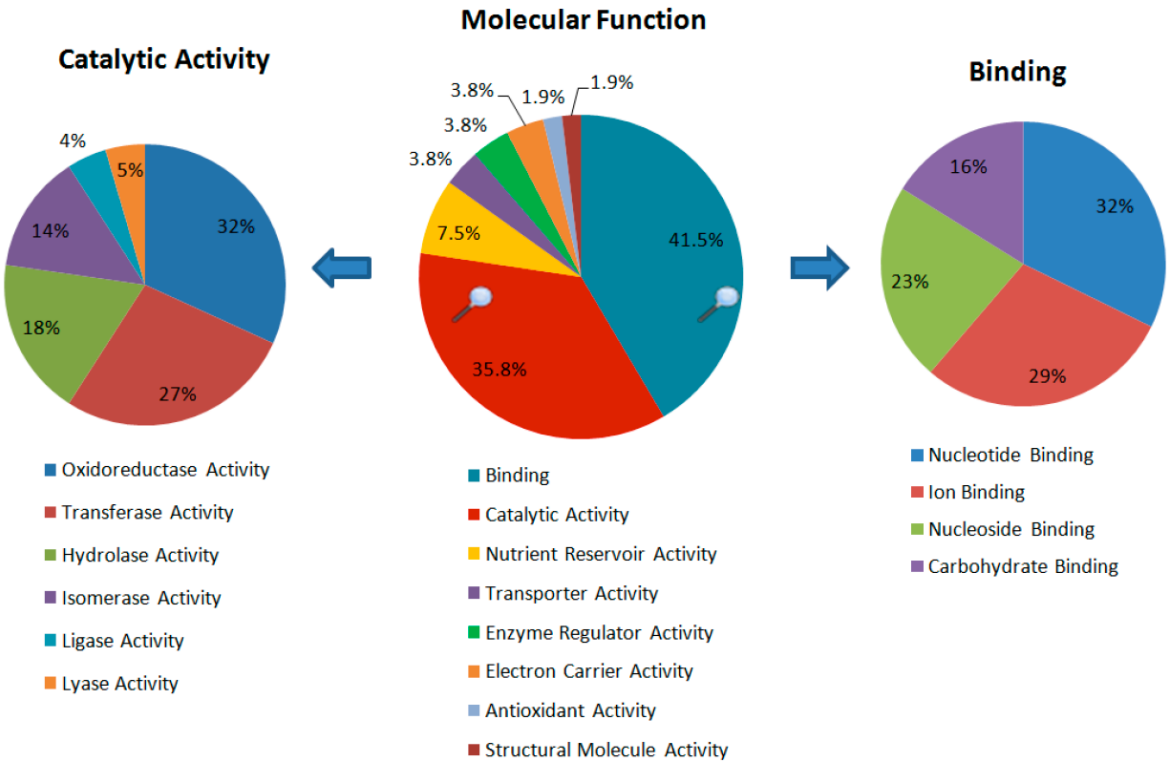

Figure 2: GO functional analysis of the proteins identified in *Phaseolus vulgaris* seeds. A total of 43 unique proteins were analyzed to which 69 GO "molecular function" terms were assigned. Pie charts are shown for the resulting 8 categories of molecular functions as well as a more detailed analysis of the most abundant molecular functions on the left and right sides.

dehydrogenase. Alcohol dehydrogenases (ADHs) are a group of dehydrogenases that facilitate the interconversion of alcohols and aldehydes or ketones. Protein spots that are involved in carbohydrate metabolism include one spot (# 56) identified as the large subunit of ribulose -1, 5-bisphosphate carboxlase/oxygenase, the highly abundant protein responsible for the initiation of the photosynthetic carbon metabolism**.** Also, UTP-glucose 1-phosphate uridylyltransferase 1 (# 19) starch synthase (# 130), and glutathione transferase (# 98) were found, which is similar to Vensel *et al*. [42] found in wheat endosperm. We found one spot (# 18) identified as tubulin β-chain, which increases during germination and may be associated with cell division. Other proteins observed in our 2D-PAGE analysis include Em protein (# 139), calreticulin (# 17), IAA protein conjugate (# 57), luminal binding protein (# 2), yieldin (# 67, 70), and F1-ATPase alpha subunit (# 16).

Protein Classification of the Identified Proteins

To better understand the functions of the identified proteins, we used the GO database [43] to categorize the functions of the common bean proteins. Agri-GO analysis toolkit [44] was used to obtain single enrichment analysis using GO terms. A total of 141 proteins were subjected to molecular functional analysis. A significant number of the identified proteins (41.5%) are classified as being involved in binding activity. About 35.8% of the identified proteins have catalytic activity, while 7.5% exhibit nutrient reservoir activity. A more detailed analysis of binding proteins showed that 32% of these proteins could bind to nucleotides. The detailed analysis of proteins classified as having catalytic activity showed that 32% of them exhibit oxidoreductase activity. Minor fractions of proteins found in our analysis exhibit antioxidant activity (1.9%), transporter activity (3.8%), enzyme regulator activity (3.8%), structural molecule activity (1.9%), or electron carrier activity (3.8%) (Figure **2**).

CONCLUSION

We separated common bean proteins using 2D-PAGE and identified 141 protein spots with MALDI-TOF/TOF tandem mass spectrometry. The proteomic characterization of the common bean seed described here will be a useful tool in understanding the physiological and molecular events that occur during seed development, and further the investigation of the seed germination process. This information will be useful to scientists wishing to improve the protein content in common beans.

ACKNOWLEDGEMENTS

We thank Dr. J. Slovin for critical review of this manuscript. Funding for this research was provided by ARS project 1275-21000-224-00D. Mention of trade name, proprietary product or vendor does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture or imply its approval to the exclusion of other products or vendors that also may be suitable.

REFERENCES

[1] Sathe SK. Dry bean protein functionality. Crit Rev Biotechnol 2002; 22: 175-23. http://dx.doi.org/10.1080/07388550290789487

[2] Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, *et al.* Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. Proc Natl Acad Sci USA 2012; 109: E788-796. http://dx.doi.org/10.1073/pnas.1108973109

[3] Balsamo GM, Cangahuala-Inocente GC, Bertoldo JB, Terenzi H, Arisi ACM. Proteomic analysis of four Brazilian MON810 maize varieties and their four non-genetically-modified isogenic varieties. J Agric Food Chem 2011; 59: 11553-59. http://dx.doi.org/10.1021/jf202635r

[4] Natarajan S, Xu C, Caperna TJ, Garrett WM. Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. Anal Biochem 2005; 342: 214-20. http://dx.doi.org/10.1016/j.ab.2005.04.046

[5] Natarajan SS, Xu C, Bae H, Caperna TJ, Garrett WM. Characterization of storage proteins in wild (Glycine soja) and cultivated (Glycine max) soybean seeds using proteomic analysis. J Agric Food Chem 2006; 54: 3114-20. http://dx.doi.org/10.1021/jf052954k

[6] Ruebelt MC, Lipp M, Reynolds TL, Astwood JD, Engel KH, Jany KD. Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of genetically modified crops. 2. Assessing natural variability. J Agric Food Chem 2006; 54: 2162-68. http://dx.doi.org/10.1021/jf052357y

[7] Louis S, Delobel B, Gressent F, Rahioui I, Quillien L, Vallier A, *et al.* Molecular and biological screening for insect-toxic seed albumins from four legume species. Plant Sci 2004; 167: 705-14. http://dx.doi.org/10.1016/j.plantsci.2004.04.018

[8] Marsolais F, Pajak A, Yin F, Taylor M, Gabriel M, Merino DM, *et al.* Proteomic analysis of common bean seed with storage protein deficiency reveals up-regulation of sulfur-rich proteins and starch and raffinose metabolic enzymes, and down-regulation of the secretory pathway. J Proteomics 2010; 73: 1587-600. http://dx.doi.org/10.1016/j.jprot.2010.03.013

[9] Osborne TB, Campbell GF. Proteids of the pea. J Am Chem Soc 1898; 20: 348-62. http://dx.doi.org/10.1021/ja02067a006

[10] Millerd A. Biochemistry of legume seed proteins. Ann Rev Plant Physiol 1975; 26: 53-72. http://dx.doi.org/10.1146/annurev.pp.26.060175.000413

[11] Tiedemann J, Neubohn B, Muntz K. Different functions of vicilin and legumin are reflected in the histopattern of globulin mobilization during germination of vetch (*Vicia sativa L.*). Planta 2000; 211: 1-12. http://dx.doi.org/10.1007/s004250000259

[12] Salmanowicz BP. Phaseolin seed variability in common bean (*Phaseolus vulgaris L.*) by capillary gel electrophoresis. J Appl Genet 2001; 42: 269-81.

[13] De La Fuente M, Borrajo A, Bermudez J, Lores M, Alonso J, Lopez M, *et al.* 2-DE-based proteomic analysis of common bean (*Phaseolus vulgaris L.*) seeds. J Proteomics 2011; 74: 262-67. http://dx.doi.org/10.1016/j.jprot.2010.10.004

[14] De La Fuente M, Lopez-Pedrouso M, Alonso J, Santalla M, De Ron AM, Alvarez G, *et al.* In-depth characterization of the Phaseolin protein diversity of common bean (*Phaseolus vulgaris L.*) based on two-dimensional electrophoresis and mass spectrometry. Food Technol Biotechnol 2012; 50: 315-25.

[15] Bollini R, Vitale A. Genetic variability in charge microheterogeneity and polypeptide compositions of *Phaseolus vulgaris*; and peptide maps of its three major subunits. Plant Physiol 1981; 52: 96-103. http://dx.doi.org/10.1111/j.1399-3054.1981.tb06040.x

[16] Montoya CA, Leterme P, Victoria NF, Toro O, Souffrant WB, Beebe S, *et al.* Susceptibility of phaseolin to *in vitro* proteolysis is highly variable across common bean varieties (*Phaseolus vulgaris*). J Agric Food Chem 2008; 56: 2183-91. http://dx.doi.org/10.1021/jf072576e

[17] Slightom JL, Drong RF, Klassy RC, Hoffman LM. Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. Nucleic Acids Res 1985; 13: 6483-98. http://dx.doi.org/10.1093/nar/13.18.6483

[18] Derbyshire E, Wright DJ, Boulter D. Legumin and vicilin, storage proteins of legume seeds. Phytochemistry 1976; 15: 3-24. http://dx.doi.org/10.1016/S0031-9422(00)89046-9

[19] Simon AE, Tenbarge KM, Scofield SR, Finkelstein RR, Crouch ML. Nucleotide sequence of a cDNA clone of Brassica napus 12S storage protein shows homology with legumin from Pisum sativum. Plant Mol Biol 1985; 5: 191-201. http://dx.doi.org/10.1007/BF00015683

[20] Mirkov TE, Wahlstrom JM, Hagiwara K, Finardi-Filho F, Kjemtrup S, Chrispeels MJ. Evolutionary relationships among proteins in the phytohemagglutinin-arcelin-alpha-amylase inhibitor family of the common bean and its relatives. Plant Mol Biol 1994; 26: 1103-13. http://dx.doi.org/10.1007/BF00040692

[21] Lis H, Sharon N. Lectins in higher plants, In: Marcus A, editors. The Biochemistry of Plants, New York, London: Academic Press 1981; pp. 371-477.

[22] Schaffrath B, Mengs U, Schwarz T, Hilgers RD, Beuth J, Mockel B, *et al.* Anticancer activity of rViscumin (recombinant mistletoe lectin) in tumor colonization models with immunocompetent mice. Anticancer Res 2001; 21: 3981-87.

[23] Sturm A, Chrispeels MJ. The high mannose oligosaccharide of phytohemagglutinin is attached to asparagine 12 and the modified oligosaccharide to asparagine 60. Plant Physiol 1986; 81: 320-22. http://dx.doi.org/10.1104/pp.81.1.320

[24] De Hoff PL, Brill LM, Hirsch AM. Plant lectins: the ties that bind in root symbiosis and plant defense. Mol Genet Genomics 2009; 282: 1-15. http://dx.doi.org/10.1007/s00438-009-0460-8

[25] Qi RF, Song ZW, Chi CW. Structural features and molecular evolution of Bowman-Birk protease inhibitors and their potential application. Acta Biochim Biophys Sin (Shanghai) 2005; 37: 283-92. http://dx.doi.org/10.1111/j.1745-7270.2005.00048.x

[26] Lioi L, Nuzzi A, Campion B, Piergiovanni A. Assessment of genetic variation in common bean (*Phaseolus vulgaris L.*)

from Nebrodi mountains (Sicily, Italy). Genetic Resources Crop Evolution 2012; 59: 455-64.
http://dx.doi.org/10.1007/s10722-011-9696-3

[27] Rosenberg LA, Rinne RW. Moisture loss as a prerequisite for seedling growth in soybean seeds (*Glycine max L. Merr*.). J Exper Botany 1986; 37: 1663-74.
http://dx.doi.org/10.1093/jxb/37.11.1663

[28] Thomashow MF. Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. Annu Rev Plant Physiol Plant Mol Biol 1999; 50: 571-99.
http://dx.doi.org/10.1146/annurev.arplant.50.1.571

[29] Grelet J, Benamar A, Teyssier E, Avelange-Macherel MH, Grunwald D, Macherel D. Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. Plant Physiol 2005; 137: 157-67.
http://dx.doi.org/10.1104/pp.104.052480

[30] Dalal M, Tayal D, Chinnusamy V, Bansal KC. Abiotic stress and ABA-inducible Group 4 LEA from Brassica napus plays a key role in salt and drought tolerance. J Biotechnol 2009; 139: 137-45.
http://dx.doi.org/10.1016/j.jbiotec.2008.09.014

[31] Liu G, Xu H, Zhang L, Zheng Y. Fe binding properties of two soybean (Glycine max L.) LEA4 proteins associated with antioxidant activity. Plant Cell Physiol 2011; 52: 994-1002.
http://dx.doi.org/10.1093/pcp/pcr052

[32] Shih MD, Hsieh TY, Lin TP, Hsing YI, Hoekstra FA. Characterization of two soybean (*Glycine max L*.) LEA IV proteins by circular dichroism and Fourier transform infrared spectrometry. Plant Cell Physiol 2010; 51: 395-407.
http://dx.doi.org/10.1093/pcp/pcq005

[33] Natarajan SS. Natural variability in abundance of prevalent soybean proteins. Regul Toxicol Pharmacol 2010; 58: S26-29.
http://dx.doi.org/10.1016/j.yrtph.2010.08.005

[34] Aghaei K, Ehsanpour AA, Shah AH, Komatsu S. Proteome analysis of soybean hypocotyl and root under salt stress. Amino Acids 2009; 36: 91-98.
http://dx.doi.org/10.1007/s00726-008-0036-7

[35] Hong SW, Vierling E. Mutants of Arabidopsis thaliana defective in the acquisition of tolerance to high temperature stress. Proc Natl Acad Sci USA 2000; 97: 4392-97.
http://dx.doi.org/10.1073/pnas.97.8.4392

[36] Wehmeyer N, Hernandez LD, Finkelstein RR, Vierling E. Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation. Plant Physiol 1996; 112: 747-57.
http://dx.doi.org/10.1104/pp.112.2.747

[37] Wang W, Vinocur B, Shoseyov O, Altman A. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci 2004; 9: 244-52.
http://dx.doi.org/10.1016/j.tplants.2004.03.006

[38] DeRocher A, Vierling E. Cytoplasmic HSP70 homologues of pea: differential expression in vegetative and embryonic organs. Plant Mol Biol 1995; 27: 441-56.
http://dx.doi.org/10.1007/BF00019312

[39] Ahsan N, Donnart T, Nouri MZ, Komatsu S. Tissue-specific defense and thermo-adaptive mechanisms of soybean seedlings under heat stress revealed by proteomic approach. J Proteome Res 2010; 9: 4189-204.
http://dx.doi.org/10.1021/pr100504j

[40] Stupnikova I, Benamar A, Tolleter D, Grelet J, Borovskii G, Dorne AJ*, et al.* Pea seed mitochondria are endowed with a remarkable tolerance to extreme physiological temperatures. Plant Physiol 2006; 140: 326-35.
http://dx.doi.org/10.1104/pp.105.073015

[41] Scandalios JG. Response of plant antioxidant defense genes to environmental stress. Adv Genet 1990; 28: 1-41.
http://dx.doi.org/10.1016/S0065-2660(08)60522-2

[42] Vensel WH, Tanaka CK, Cai N, Wong JH, Buchanan BB, Hurkman WJ. Developmental changes in the metabolic protein profiles of wheat endosperm. Proteomics 2005; 5: 1594-11.
http://dx.doi.org/10.1002/pmic.200401034

[43] Dimmer EC, Huntley RP, Barrell DG, Binns D, Draghici S, Camon EB*, et al.* The Gene Ontology — providing a functional role in proteomic studies. Proteomics 2008; 8: n/a-n/a.

[44] Du Z, Zhou X, Ling Y, Zhang Z, Su Z. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 2010; 38: W64-70.
http://dx.doi.org/10.1093/nar/gkq310

Received on 29-04-2013 Accepted on 11-06-2013 Published on 17-06-2013

http://dx.doi.org/10.6000/1927-5129.2013.09.55