

A Systematic Proteome Study of Seed Storage Proteins from Two Soybean Genotypes

Seong-Woo Cho^{***}, Soo-Jeong Kwon^{*}, Swapan Kumar Roy^{*}, Hong-Sig Kim^{*}, Chul-Won Lee^{*}, and Sun Hee Woo^{*†}

^{*}Department of Crop Science, Chungbuk National University, Cheong-ju 361-763, Korea;

^{**}Rice Research Division, National Institute of Crop Science, RDA, Suwon 441-857, Korea

ABSTRACT Soybean seed is a good source of plant protein in human consumables such as baby formula and protein concentrate. The seeds contain an abundance of storage proteins, namely β -conglycin and glycinin that account for ~ 70-80% of the total seed protein content. Proteome profiling has been proved to be an efficient way that can help us to investigate the seed storage proteins. In the present study, the seeds were removed from the pods and the cotyledonary tissues were separated from the testa for proteome analysis in order to investigate the seed storage proteins. A systematic proteome profiling was conducted through one-dimensional gel electrophoresis followed by MALDI-TOF-TOF mass spectrometry in the seeds (cotyledonary tissue) of soybean genotypes. Two dimensional gels stained with CBB, a total of 10 proteins were identified and analyzed using MASCOT search engine according to the similarity of sequences with previously characterized proteins along with the UniProt database. A total of ten proteins such as glycinin Gy4 precursor, glycinin G3 precursor, glycinin G1 precursor, glycinin chain A2B1a precursor, glycinin chain A2B1a precursor were identified in our investigation. However, the glycinin subunit may be considered to play important roles in soybean breeding and biochemical characterization. In addition, the improved technique will be useful to dissect the genetic control of glycinin expression in soybean.

Keywords : soybean, proteome analysis, seed storage proteins, glycinin subunit, cotyledon

Seed crops are basically cultivated for the stored protein, oil, and carbohydrates that specifically accumulate in seeds. Seeds can be regarded as the storing primarily either protein and triglyceride or protein and carbohydrate (Schmidt *et al.*, 2011). Soybean (*Glycine max* L. Merr.) seeds contain proteins in a concentration of 40-45% on a dry-weight basis. Glycinin,

a major seed storage protein, accounts for about 35% of total seed protein. The relative abundance of glycinin, a major reserve protein found in soybean seed, has prompted and facilitated the study of its structure by numerous workers. Glycinin consists of six subunits, each made up of an acidic and a basic polypeptide component linked by a single disulfide bond (Staswick *et al.*, 1984). Native glycinin contains 6 subunits and has a molecular weight of approximately 350,000 and five basic components which were identified unambiguously on the basis of differences in the partial primary structure of their *N*-terminal regions (Ereken-Tumer *et al.*, 1982).

Research revealed that seed storage proteins of Soybean (*Glycine max*) are encoded by a few conserved gene families (Schuler *et al.*, 1982a, 1982b; Harada *et al.*, 1989; Nielsen *et al.*, 1989). Most seed proteins belongs to the members of the cupin superfamily (e.g. legumins and vicilins), but in some dicotyledonous seeds, lectins are abundant, whereas in cereal grains, the prolamins and to a lesser extent the legumins are abundant (Herman and Larkins, 1999).

Each subunit complexes consists of an acidic polypeptide component which is disulfide-linked to a basic polypeptide component. These acidic subunits were designated A1a, A1b, A2, A3, A4, A5 and A6, while the basic ones were designated B1a, B1b, B2, B3 and B4 (Staswick and Nielsen, 1983). The complete amino acid sequences of all of the acidic subunits except the A1b and A6 subunits have been determined by protein sequencing of deduced by DNA sequence analysis (Hirano *et al.*, 1984).

Five major subunits have been grouped into two subunit groups based on sequence homology, i.e. group I (A1aB2, A1bB1b, and A1B1a) and group II (A3B4 and A5A4B3) (Nielsen, 1985). It is known that about 20% Japanese soybean varieties lack the A5A4B3 subunit (Harada *et al.*, 1983). A

[†]Corresponding author: (Phone) +82-43-261-2515 (E-mail) shwoo@chungbuk.ac.kr

<Received 19 August, 2014; Accepted 25 August, 2014>

mutant soybean line lacking group I subunits was induced by r-ray irradiation (Odanaka and Kaizuma, 1987).

Recently, in the field of plant proteomics, plant biology has been increasing interest in the proteins expressed by the genome, the proteome. In this context, two-dimensional gel electrophoresis is a powerful tool to analyze complex protein mixtures by separating the proteins according to their isoelectric point (*pI*) and their molecular mass, as first described by O'Farrell (1975). Although these technologies had been improved by the development of the better carrier ampholyte production, acrylamide polymers for isoelectric focusing, and experimental instrumentation, probably, the single most important advance in isoelectric focusing for proteome analysis was the development of immobilized pH gradient (IPG) technology. Also, fingerprint is obtained with the aid of high-resolution techniques such as two-dimensional electrophoresis.

To perform proteome analysis 2-DE technique is very important because a number of proteins are separated by mainly two-dimensional electrophoresis (2-DE). However, glycinin subunit is thought to be one of the most important elements that are also considered to play important roles in soybean breeding and biochemical characterization. To clarify how glycinin subunit affects in soybean seeds, we studied the seed storage proteins.

MATERIALS AND METHODS

Plant materials and sample preparation

Wild type soybean genotypes (IT 184222, IT 188365, IT 183033) and Korean cultivar, Hwangkeumkong were used for this investigation. The seeds were aseptically removed from the pods and the cotyledonary tissues were separated from the testa and the embryo were frozen in liquid N₂ to retain polysome integrity. Solubilization mix was added to the sample equally. The suspension was briefly sonicated and allowed to extract for 1 hr at room temperature. A microsomal fraction was obtained from a cotyledon homogenate and the resulting supernatant was used for this experiments. A portion (4g) of fresh cotyledons were homogenized with 2 mL of phosphate buffer pH 7.6 containing 65 mM K₂HPO₄, 2.6 mM KH₂PO₄, 400 mM NaCl and 3 mM NaN₃ at 4°C. Proteins were solubilized with 2 µg/10µL SDS and precipitated with Cytocrome c, Riso

and Histone.

Two-dimensional gel electrophoresis

The high throughput two-dimensional gel electrophoresis was carried out according to the protocol of O'Farrell (1975). Sample solutions (400µg) were loaded onto the acidic side of the IEF tube gels (11 cm x 2 mm), which was pre-run at 150 V for 1 h, 300 V for 1 h, and 500 V for 16 h for the first dimension. In order to avoid the overlapping of protein spots and to enhance the resolution capacity, an IEF gel was adopted specifically for pH range 3 to 10 (carrier ampholyte) in addition to the acidic and basic ranges. SDS-PAGE in the second dimension (Nihon Eido, Tokyo, Japan) was executed with 12 % separation and 5 % stacking gels with 13 cm x 13 cm gel plates. Protein spots in the 2-DE gels were visualized by Coomassie Brilliant Blue (CBB R-250)-staining. Each biological sample was carried out three times and the visualized gels were selected for image analysis in each replication.

Protein visualization and image analysis

After staining, gels were scanned using a flatbed scanner, and the data were analyzed using PDQuest software version 6.1 (Bio-Rad). After selecting so-called landmarks and the assignment of all features, two dimensional electrophoresis images were aligned and matched. Three independent biological replications were carried out to validate the results. Isoelectric points (*PI*) and relative molecular weights (*Mr*) of proteins were inferred using *pI* (2-D standard; Bio-Rad Laboratories, Richmond, CA, USA) and *Mr* (LMW marker kit; Amersham Pharmacia Biotech) marker proteins as standards.

In-gel digestion

CBB-stained gel slices washed with 30% methanol until the colors were completely removed. Then the gel slices were destained with 10mM (NH₄) HCO₃ in 50% ACN (Acetonitrile), squeezed for 10min with 100% ACN (Acetonitrile) and dried by vacuum centrifugation. After destaining steps, the gel slices were reduced with 10mM DTT in 100mM (NH₄)HCO₃ at 56°C for 1hr and then alkylated with 55mM Iodoacetamide (IAA) in 100mM (NH₄)HCO₃ in the dark for 40min. Then the gel slices were digested with 50 µL trypsin buffer (Promega Corporation, Madison, WI 53711-5399, USA) and incubated at 37°C for 16hr. After digestion steps, the peptides were

extracted with 50mM ABC (Ammonium Bi-Carbonate) and repeated this steps several times with a solution containing 0.1% formic acid in 50% ACN (Acetonitrile) until 200~250ul. The solution containing eluted peptides was concentrated up to drying by vacuum centrifugation and the resultant extracts were confirmed by MALDI-TOF-TOF mass spectrometry.

Mass spectrometry (MALDI-TOF-TOF/MS) analysis

Selected spots from 2-DE gels were analyzed to evaluate the compatibility between protein extraction and mass spectrometry as well as to disclose the protein classes that populate each 2-DE gels. Mass spectra were acquired in an ABI 4700 Proteomics Analyzer (Applied Biosystems) using 3, 5-dimethoxy-4-hydroxycinnamic acid as matrix and the resulting data by the GPS Explorer package (Applied Biosystems). Peptides were dissolved in 0.5 % (v/v) trifluoroacetic acid (TFA) and desalted with a ZipTip C18 (Millipore, Bedford, MA, USA). Those purified peptides were then eluted directly onto a MALDI plate by using an α -cyano-4- hydroxycinnamic acid (CHCA) matrix solution [10 mg per cm³ of CHCA in 0.5 % (v/v) TFA + 50 % (v/v) acetonitrile; 1:1]. All mass spectra were acquired in the reflection mode with 0 - 4000 m/z by a 4700 proteomics analyzer (Applied Bio-systems, Framingham, MA, USA). External calibration was performed using a standard peptide mixture of des-Arg bradykinin, angiotensin, Glufibrinopeptide B, adrenocorticotrophic hormone (ACTH) clip 1-17, ACTH clip 18-39, and ACTH clip 7-38.

Bioinformatics

To identify the peptides, acquired MS/MS spectra were evaluated using Mascot Generic File (MGF) with an in-house licensed MASCOT search engine (Mascot v. 2.3.01, Matrix Science, London, UK) against the viridiplantae within the NCBI nr database. MASCOT was used with the monoisotopic mass selected, a peptide mass tolerance of 50 ppm, and a fragment ion mass tolerance of 2 Da. The instrument setting was specified as MALDI-TOF/TOF. The carbamidomethylation of cysteines was set as a fixed modification whereas the oxidation of methionines was set as a variable modification. Trypsin was specified as the proteolytic enzyme with one potential missed cleavage. All proteins identified by high-scoring peptides were considered true matches, and at least two peptide matches. Protein hits were validated if the

identification involved at least 10 top-ranking peptides with $P < 0.05$ and peptide scores > 34 , and also selected false positive rate < 0.05 . When those peptides matched multiple members of a protein family, the presented protein was selected based on the highest score and the greatest number of matching peptides.

RESULTS AND DISCUSSION

Optimization of basic proteins by 2-D gel electrophoresis

Cotyledon proteins were identified from wild types soybean genotypes (*Glycine soja* L.), (IT 184222; IT 188365; IT 183033), and Korean soybean (*Glycine max* L.) cv. Hwangkeumkong; which is an early ripening variety and takes about 60 days after flowering to mature its seeds. A master 2-D gel electrophoresis with an isoelectric focusing range of pH 3.5-10 and a potential load of 50 μ L were performed and resulting in a gel with more than 300 features visualized by CBB staining (Fig. 1). Most cotyledon polypeptides had a isoelectric point between 5.0-7.0. Proteins were automatically identified on two-dimensional gels electrophoresis by PDQuest software based on molecular weights and isoelectric points. The two-dimensional electrophoresis images were aligned and matched.

Identification of subunit of the soybean cotyledon proteins

We attempted to separate the basic proteins by two-dimensional gels electrophoresis that may affect the reduction of the number sequences in soybean cotyledon proteins. The new sequences are 100% of the sequenced 10 proteins. Among the sequenced 10 proteins, proteins (50%) were new proteins which listed in PIR, EMBL and SWISS-PROT databases. The relatively low identification ratio and higher percentage of new sequence of *Arabidopsis* and maize proteins as compared with soybean proteins may reflect the partially their genomic and cDNA sequencing work in soybean. The sequence comparisons showed sixteen proteins characterized in soybean with 88% to 100% identity (Table 1): glycinin (two spots, M 87064, L 04967). In plants, cotyledon proteins, a large proportion of novel or functionally uncharacterized proteins were found. Very good probabilities were given by the FASTA or BLAST soft program for this identification.

Glycinin, a hexameric 11S globulin seed storage protein

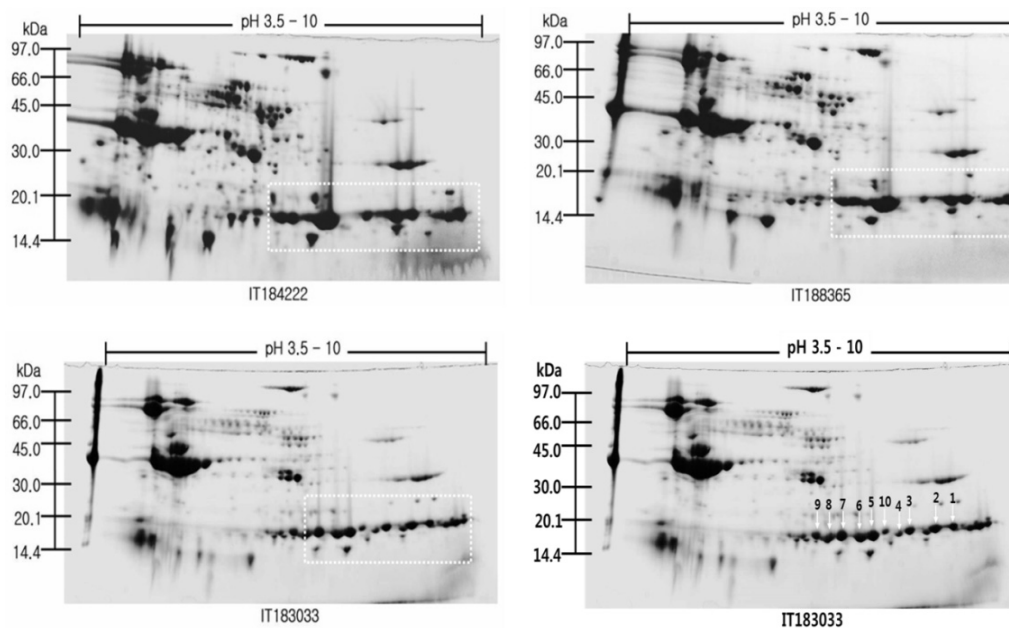


Fig. 1. Reference 2-DE gel map from the Hwangkeumkong cultivar of Soybean. Proteins were extracted from different organs, separated by 2-DE, and visualized by CBB. The MW of each protein was determined by 2-DE markers.

Table 1. List of identified seed proteins from wild type cultivars and Korean cultivar using MALDI-TOF-TOF mass spectrometry.

No. of spot ^z	UniGene contig	Total MOWSE score	pI	Mr	Identify	Genbank accession	Protein
1	S20946	326	5.21	63.8	100%	gi 81779	Glycinin Gy4 precursor
2	S20946	461	5.21	63.8	100%	gi 81779	Glycinin Gy4 precursor
3	S20946	442	5.21	63.8	100%	gi 81779	Glycinin Gy4 precursor
4	S11003	506	5.28	54.3	100%	gi 99909	Glycinin G3 precursor
5	P04776*	421	5.89	55.6	100%	gi 121276	Glycinin G1 precursor
6	P04405*	321	5.56	54.2	100%	gi 72295	Glycinin chain A2B1a precursor
7	P04776*	329	5.89	55.6	100%	gi 121276	Glycinin G1 precursor
8	P04405*	349	5.56	54.2	100%	gi 72295	Glycinin chain A2B1a precursor
9	P04776*	228	5.89	55.6	100%	gi 121276	Glycinin G1 precursor
10	S20946	315	5.21	63.8	100%	gi 81779	Glycinin Gy4 precursor

indicates SWISS-PROT accession number

^zindicates that numbers of spots are shown at Fig. 4

(360 kDa), consists of both acidic (A) and basic (B) polypeptides. Based on physical properties, these five subunits are classified into two distinct major groups. Previous results revealed that glycinin is composed of five subunits, G1, G2, G3, G4, and G5, the precursors of which are encoded by five non-allelic genes, Gy1, Gy2, Gy3, Gy4, and Gy5, respectively (Nielsen *et al.*, 1989). Various subunits of glycinin are considered to play

different important roles in tofu gel formation. The presence of more G4 subunit accelerates the gel formation rate and transparency, whereas the presence of more G5 is related to gel hardness. In addition, glycinin is thought to be one of the most important storage proteins that reduces cholesterol levels in human serum. There are numerous reports on molecular characterization of soybean seeds at the protein and DNA

levels which are crucial for assessing genetic diversity of different genotypes (Natarajan *et al.*, 2007). However, in the present work, four spots (1, 2, 3, 10) were identified as glycinin Gy4 precursor, three protein spots (5, 7, 9) were identified as glycinin G1 precursor, one protein spot (6) was identified as glycinin chain A2B1a precursor and one protein spot (8) was identified as glycinin chain A2B1a precursor.

The usual tools were used to query the databases (FASTA and BLAST). For identified spots, an average of match of 45% identify, over an average sequence length of 10 amino acid, was found. The probability that these obtained scores is low, but two points deserve further discussion for one-third of the identified spots. This is particularly likely for spots matching with partly sequenced ESTs. The incomplete matches may also mean that these spots correspond to isoforms not present in the databases. However, it can't be ruled out that different proteins co-migrated in the same spot. Another striking feature is that different spots matched with the same protein type. However, the relative molecular weight measured on gels is lower than that calculated for the corresponding translated protein (by less than 4) for only two spots. This does not preclude any degradation but suggests that, it would concern a minor proportion of identified proteins. Therefore, the presence of multiple spots with similar sequences probably reflects the presence of isoforms and/or of post-translational modifications.

ACKNOWLEDGMENTS

This work was supported by the research grant of the Chungbuk National University in 2012.

REFERENCES

- Ereken-Tumer, N., J. D. Richer, and N. C. Nielsen. 1982. *J. Biol. Chem.* 257 : 4016-4018.
- Harada J. J., S. J. Barker, and R. B. Goldberg. 1989. Soybean β -conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and post-transcriptional processes. *Plant Cell* 1 : 415-425.
- Harada, K., Y. Toyokawa, and K. Kitamura. 1983. Genetic analysis of the most acidic 11S globulin subunit and related characters in soybean seeds. *Jpn. J. Breed.* 33(1) : 23-33.
- Herman E. M. and B. A. Larkins. 1999. Protein storage bodies and vacuoles. *Plant Cell* 11 : 601-614.
- Hirano, T. O. S. H. I. O., K. O. U. J. I. Fujimoto, T. S. U. Y. O. S. H. I. Teranishi, N. A. O. K. I. Nishino, K. Onoue, S. Maeda, and K. Shimada. 1984. Phorbol ester increases the level of interleukin 2 mRNA in mitogen-stimulated human lymphocytes. *J. Immunol.* 132(5) : 2165-2167.
- Natarajan, S. S., H. B. Krishnan, F. Khan, X. Chen, W. M. Garrett, and D. Lakshman. 2013. Analysis of soybean embryonic axis proteins by two-dimensional gel electrophoresis and mass spectrometry. *J. Basic Applied Sci.* 9 : 302-308.
- Natarajan, S., C. Xu, H. Bae, B. A. Bailey, P. Cregan, Caperna, J. C. Thomas, M. G. Wesley, and L. Devanand. 2007. Proteomic and genetic analysis of glycinin subunits of sixteen soybean genotypes. *Plant physiol. Biochem.* 45(6) : 436-444.
- Nielsen N. C., C. D. Dickinson, T. J. Cho, V. H. Thanh, and B. J. Scallan. 1989. Characterization of the glycinin gene family in soybean. *Plant Cell* 1 : 313-328.
- Nielsen N. C., C. D. Dickinson, T. J. Cho, V. H. Thanh, B. J. Scallan, R. L. Fischer, T. L. Sims, G. N. Drews, and R. B. Goldberg. 1989. Characterization of the glycinin gene family in soybean. *Plant Cell* 1 : 313-328.
- Nielson, N. C. 1985. The structure and complexity of the 11S polypeptide in soybean. *J. Am. Oil Chem.* 49 : 2733-2740.
- O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250 : 4007-4021.
- Odanaka, H. and N. Kaizuma. 1987. Mutants on soybean storage protein induced with r-ray irradiation (in Japanese). *Jpn. J. Breed.* 39(Suppl. 1) : 430-431.
- Schmidt, M. A., W. B. Barbazuk, M. Sandford, G. May, Z. Song, W. Zhou, and E. M. Herman. 2011. Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. *Plant physiol.* 156(1) : 330-345.
- Schuler M. A., B. F. Ladin, G. Fryer, J. G. Pollaco, and R. N. Beachy. 1982a. Structural sequences are conserved in the genes coding for the alpha, alpha' and beta-subunits of the soybean seed storage protein and other seed proteins. *Nucleic Acids Res.* 10 : 8245-8260.
- Schuler M. A., E. S. Schmitt, and R. N. Beachy. 1982b. Closely related families of genes code for the alpha and alpha' subunits of the soybean 7S storage protein complex. *Nucleic Acids Res.* 10 : 8225-8244.
- Staswick, P. E., P. Broue, and N. C. Nielsen. 1983. Glycinin composition of several perennial species related to soybean. *Plant physiol.* 72(4) : 1114-1118.
- Staswick, P. E., M. A. Hermodson, and N. C. Nielsen. 1984. Identification of the cystines which link the acidic and basic components of the glycinin subunit. *J. Biol. Chem.* 259 : 13431-13435.