

Analysis of Soybean Seed Proteins Using Proteomics

Savithiry S Natarajan*

U.S. Department of Agriculture, Agricultural Research Service, Soybean Genomics and Improvement Laboratory, Beltsville, MD 20705, USA

Soybean food products are popular because of their health benefits and they are a major source of protein for various food supplements. Soybean seed contains 40-50% protein on a dry matter basis, making it 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, which account for ~70-80% of the total seed protein content. Lesser abundant proteins include β -amylase, cytochrome c, lectin, lipoxygenase, urease, Kunitz trypsin inhibitor (KTI), and the Bowman Birk inhibitor (BBI) of chymotrypsin and trypsin [1].

In order to determine the variation of seed proteins that may occur in the crop as a result of genetic modification, accurate and replicable methodology for protein extraction, isolation, and characterization is essential. Modern proteomic tools are being used to study the expression of proteins to examine alterations in protein profiles caused by genetic mutations and environmental stress [2-10]. Proteome analysis is performed using a variety of methods including structural proteomics such as high throughput (HT) X-ray crystallography and HT nuclear magnetic resonance (NMR) spectroscopy; expressional or analytical proteomics such as gel based electrophoresis (IDE, 2DE, 2DIGE), gel-free (LC-MS/MS or multidimensional protein identification technology (Mud PIT), protein chips, DNA chips, mass spectrometry (MS), micro sequencing; and functional or interaction proteomics such as HT functional assays, ligand chips, yeast 2-hybrid, deletion analysis, and motif analysis [11-14]. In this mini review, we discuss some of the expression analysis methodology that we've applied to study soybean seed proteins. Because the extraction of soybean seed proteins for accurate 2-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is challenging, we initially optimized protein extraction techniques by comparing different methods. Extraction of protein suitable for 2D-PAGE is sample-dependent and is achieved by optimizing the concentration of chaotropic agents, detergents, reducing agents, buffers, enzymes and ampholytes. Also helpful in this regard are advances in immobile pH gradient (IPG) technology and the development of electrophoretic instruments have improved the reproducibility of protein separation. The availability of commercial IPG strips in linear and nonlinear gradients with multiple narrow pH ranges [15], allows effective protein separation for various downstream analyses. We compared four different solubilization methods (urea/thiourea, urea, modified trichloroacetic acid (TCA)/acetone and phenol) for extraction of proteins from soybean seeds for subsequent analysis by 2D-PAGE [16]. Our study demonstrated that the modified TCA/acetone method with urea/thiourea solubilization resolved more protein spots than the urea or phenol extractions. Using the phenol/urea method, resolution of proteins was generally poor and spots diffuse in the high molecular weight region of the gel, particularly when separating the proteins at pH 4.0 to 7.0. Moreover, while overall protein separation was similar in the TCA/acetone and urea/thiourea methods, low molecular weight proteins were more consistently resolved with the TCA/acetone method. In addition, characterization of low abundant proteins is a challenge because they are often masked by highly abundant proteins. Recently, Boschetti and Righetti [17] published a review of plant proteomic methods to isolate low abundance proteins. In the past, we also developed methodologies

that effectively remove the highly abundant proteins in soybean seed enabling detection of low abundant proteins previously unknown [18,19]. To improve resolution and protein identification, we used both wide (pH 3-10) and narrow (pH 4-7 and 6-11) pH gradients to separate the seed proteins that include storage, allergen, and anti-nutritional proteins (Figure 1). Proteins were then extracted from small gel pieces and analyzed by MALDI-TOF-MS, LC-MS/MS and identified by searches using the NCBI non-redundant protein database [20-22]. Based on the proteins we identified, we developed a comprehensive seed protein database named SoyProDB [23].

The soybean storage proteins, β -conglycinin and glycinin, are classified based on their sedimentation coefficients. β -conglycinins are encoded by two mRNA groups. The first mRNA group encodes α and α' β -conglycinin subunits and the second mRNA group encodes the β -subunit of β -conglycinin [24,25]. To adequately resolve the multiple protein subgroups, we used narrow pH strips, 4.0-7.0 for acidic proteins and pH 6.0-11.0 range for basic proteins. This approach resolved 7 α subunit protein spots (#1-7), an α' subunit of β -conglycinin (#8) and 6 β -subunits of β -conglycinin (#9-14). Using these methods, we reported the variation of the above proteins in wild and cultivated soybean genotypes [21]. Schuler et al. [25] reported that β -conglycinin subunits are products of a multigene family. They suggested that the variation in the distribution of protein spots may be caused by post-translational modifications [24,26]. 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 [27]. To effectively separate glycinin proteins, we used pH 3.0-10.0, 4.0-7.0 and 6.0-11.0 to resolve acidic and basic chains of glycinin subunits. The G1 subunit showed 3 basic polypeptides (spot #15-17); G2 subunit showed 8 spots (#18-25) of acidic and basic polypeptides; G3 subunit showed 8 spots (#26-33); G4 subunit showed 12 spots (#34-40); G5 subunit showed 7 seven spots (#41-47). The relative amount of all aforementioned glycinin subunits showed variation among 16 soybean genotypes examined [28].

Soybean seed also contains several allergen proteins. Gly m Bd 60K has been described as storage seed proteins as well as a major allergen in soybean. Gly m Bd 60 K includes β -conglycinin and glycinin. Only the α subunit of β -conglycinin is reported to be allergenic for consumers [29]. The acidic polypeptides of G1 and all G2 subunits of glycinin are also reported to be allergenic [30]. However, using immunoblot

*Corresponding author: Savithiry S Natarajan, U.S. Department of Agriculture, Agricultural Research Service, Soybean Genomics and Improvement Laboratory, Beltsville, MD 20705, USA, Tel: 1-301-504-5258; E-mail: Savi.natarajan@ars.usda.gov

Received February 05, 2014; Accepted February 07 2014; Published February 09, 2014

Citation: Natarajan SS (2014) Analysis of Soybean Seed Proteins Using Proteomics. J Data Mining Genomics Proteomics 5: e113. doi:10.4172/2153-0602.1000e113

Copyright: © 2014 Natarajan SS. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

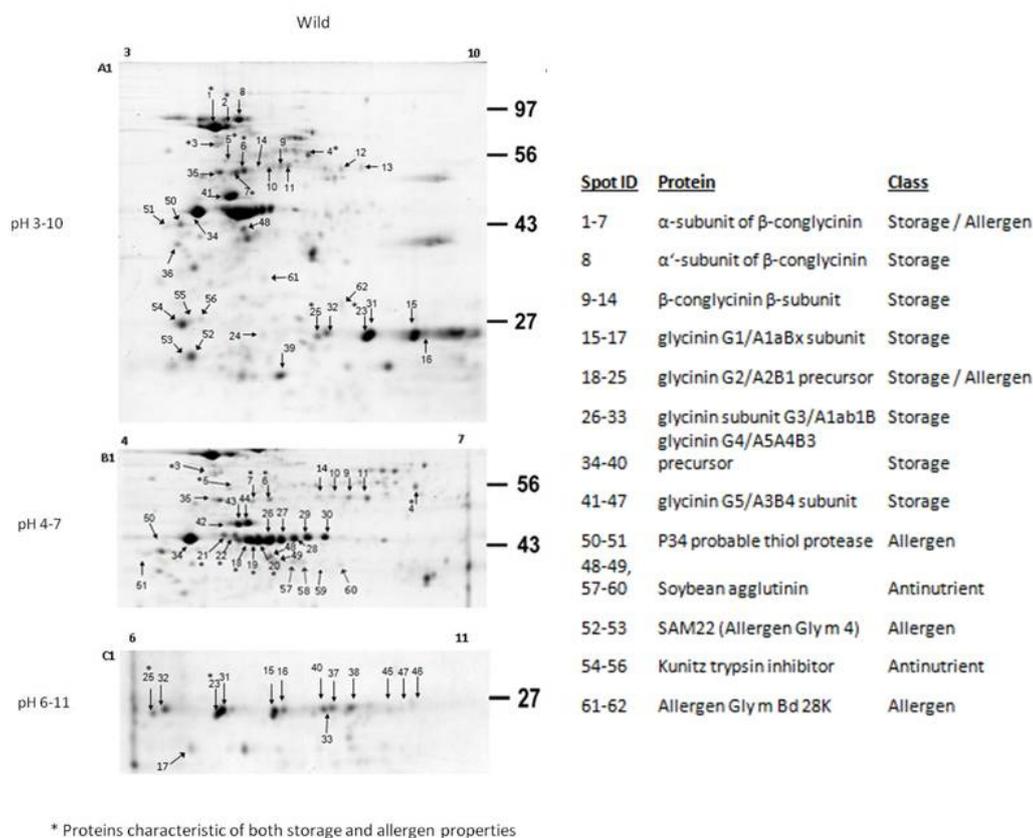


Figure 1: 2D-PAGE analysis of seed proteins in wild soybean seeds. The first dimension was run using pH gradients 3-10, 4-7, and 6-11, and the second dimension was separated by 12% SDS-PAGE.

and MALDI-TOF analysis, Krishnan et al. [18] reported that all three subunits of soybean β -conglycinin are potential food allergens. Our investigation using 2D-PAGE showed seven spots (# 1-7) of α subunits [23] as shown in Figure 1. The glycinin G2 allergens showed eight spots (#18-25). Gly m Bd 30K is another protein of soybean designated as a major allergen. This protein was previously described as the 34-kD vacuolar protein P34 [31]. Using a monoclonal antibody against P34, Yaklich et al. [32] found that P34 showed genetic diversity in soybean with respect to protein quantity. In our analyses, we reported 2 protein spots of P34 (#50, 51) in wild soybean genotypes. Gly m Bd 28 K is a less abundant allergen protein of soybean that was originally isolated from soybean meal [33]. In our study, two spots (#61, 62) of Gly m Bd 28 K were identified. A study by Xiang et al. [34] suggests that this allergen is probably processed into two smaller polypeptides of 240 and 212 amino acids in the soybean seed, and both portions are expected to be present in soybean-derived foods. These two spots of Gly m Bd 28 K probably come from the post-translational processing of the same gene.

Kunitz trypsin inhibitor (KTI) is one of the abundant anti-nutritional proteins in soybean seed that inhibits trypsin, an important animal digestive enzyme [35]. In addition, KTIs have been characterized as food allergens in humans [36]. In soybean, three KTI genes (KTI1, KTI2 and KTI3) have been reported, and both transcriptional and post-transcriptional processes regulate KTI gene expression [37]. The KTI3 transcript was detected only in the soybean seed, while KTI1 and KTI2 transcripts are expressed in soybean leaf, root, and stem. Our study showed that although the overall distribution patterns of KTI protein

spots are quite similar in wild and cultivated genotypes, the number of protein spots and their intensities varied between these genotypes [20]. Three protein spots (#54-56) of KTI were identified by MALDI-TOF-MS with database searches in wild soybean varieties. Soybean agglutinins are another class of anti-nutritional proteins present abundantly in soybeans. The presence of this protein accounts for about 10% of total protein in some legumes. The wild soybeans showed 6 spots (#48-49, 57-60) of soybean agglutinins. Based on the proteins identified in our laboratory, we developed the database SoyProDB containing comprehensive details of soybean proteins.

In conclusion, effective extraction methodology is important for analysis of both abundant and low abundant soybean seed proteins. The combination for 2D-PAGE with MS effectively differentiates several classes of soybean seed proteins. We have characterized storage, allergen, and anti-nutritional proteins among different soybean genotypes using proteomic tools. The database, SoyProDB, is useful for scientists wishing to tailor products with specific proteins to produce value-added soybean genotypes. These tools and methods will be used in ongoing investigations to study the biosafety of transgenic soybeans in our laboratory.

Acknowledgements

Funding for this research was provided by ARS project 1245-21220-232-00D.

References

1. Friedman M, Brandon DL (2001) Nutritional and health benefits of soy proteins. J Agric Food Chem 49: 1069-1086.
2. Dam S, Dyrland TF, Ussatjuk A, Jochimsen B, Nielsen K, et al. (2013) Proteome

- reference maps of the *Lotus japonicus* nodule and root. *Proteomics*.
3. Ghaffari A, Gharechahi J, Nakhoda B, Salekdeh GH (2014) Physiology and proteome responses of two contrasting rice mutants and their wild type parent under salt stress conditions at the vegetative stage. *J Plant Physiol* 171: 31-44.
 4. Hossain Z, Komatsu S (2014) Soybean proteomics. *Methods Mol Biol* 1072: 315-331.
 5. Komatsu S, Nanjo Y, Nishimura M (2013) Proteomic analysis of the flooding tolerance mechanism in mutant soybean. *J Proteomics* 79: 231-250.
 6. Lee H, Garrett WM, Sullivan JH, Forseth I, Natarajan SS (2013) Differentially expressed proteins of soybean (*Glycine max*) pulvinus in light and dark conditions. *J Basic Appl Sci* 9: 157-171.
 7. Mohammadi PP, Moieni A, Hiraga S, Komatsu S (2012) Organ-specific proteomic analysis of drought-stressed soybean seedlings. *J Proteomics* 75: 1906-1923.
 8. Nanjo Y, Nakamura T, Komatsu S (2013) Identification of indicator proteins associated with flooding injury in soybean seedlings using label-free quantitative proteomics. *J Proteome Res* 12: 4785-4798.
 9. Natarajan S, Luthria D, Bae H, Lakshman D, Mitra A (2013) Transgenic soybeans and soybean protein analysis: an overview. *J Agric Food Chem* 61: 11736-11743.
 10. Natarajan SS, Krishnan HB, Khan FH, Chen X, Garrett WM, et al. (2013) Analysis of soybean embryonic axis proteins by two-dimensional gel electrophoresis and mass spectrometry. *Journal of Basic and Applied Sciences* 9: 309-332.
 11. Bensimon A, Heck AJ, Aebersold R (2012) Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem* 81: 379-405.
 12. Kim ST, Kim SG, Agrawal GK, Kikuchi S, Rakwal R (2013) Rice proteomics: A model system for crop improvement and food security. *Proteomics*.
 13. Law KP, Lim YP (2013) Recent advances in mass spectrometry: data independent analysis and hyper reaction monitoring. *Expert Rev Proteomics* 10: 551-566.
 14. Suter B, Kittanakom S, Stagijar I (2008) Two-hybrid technologies in proteomics research. *Curr Opin Biotechnol* 19: 316-323.
 15. Görg A, Weiss W, Dunn MJ (2004) Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 4: 3665-3685.
 16. Natarajan S, Xu C, Caperna TJ, Garrett WM (2005) Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. *Anal Biochem* 342: 214-220.
 17. Boschetti E, Righetti PG (2014) Plant proteomics methods to reach low-abundance proteins. *Methods Mol Biol* 1072: 111-129.
 18. Krishnan HB, Oehrle NW, Natarajan SS (2009) A rapid and simple procedure for the depletion of abundant storage proteins from legume seeds to advance proteome analysis: a case study using *Glycine max*. *Proteomics* 9: 3174-3188.
 19. Natarajan SS, Krishnan HB, Lakshman S, Garrett WM (2009) An efficient extraction method to enhance analysis of low abundant proteins from soybean seed. *Anal Biochem* 394: 259-268.
 20. Natarajan S, Xu C, Bae H, Bailey BA (2007) Proteomic and genomic characterization of Kunitz trypsin inhibitors in wild and cultivated soybean genotypes. *J Plant Physiol* 164: 756-763.
 21. Natarajan SS, Xu C, Bae H, Caperna TJ, Garrett WM (2006) Characterization of storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using proteomic analysis. *J Agric Food Chem* 54: 3114-3120.
 22. Xu CP, Caperna TJ, Garrett WM, Cregan P, Bae HH, et al. (2007) Proteomic analysis of the distribution of the major seed allergens in wild, landrace, ancestral, and modern soybean genotypes. *J Sci Food Agr* 87: 2511-2518.
 23. Tavakolan M, Alkharouf NW, Khan FH, Natarajan S (2013) SoyProDB: A database for the identification of soybean seed proteins. *Bioinformatics* 9: 165-167.
 24. Koshiyama I (1983) Storage proteins of soybean: Seed Proteins: Biochemistry, Genetics, Nutritive Value. Springer, Netherlands.
 25. Schuler MA, Schmitt ES, Beachy RN (1982) 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.
 26. Davies CS, Coates JB, Nielsen NC (1985) Inheritance and biochemical analysis of four electrophoretic variants of β -conglycinin from soybean. *Theor Appl Genet* 71: 351-358.
 27. Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallan BJ, et al. (1989) Characterization of the glycinin gene family in soybean. *Plant Cell* 1: 313-328.
 28. Natarajan S, Xu C, Bae H, Bailey BA, Cregan P, et al. (2007) Proteomic and genetic analysis of glycinin subunits of sixteen soybean genotypes. *Plant Physiol Biochem* 45: 436-444.
 29. Wilson S, Blaschek K, de Mejia E (2005) Allergenic proteins in soybean: processing and reduction of P34 allergenicity. *Nutr Rev* 63: 47-58.
 30. Beardslee TA, Zeece MG, Sarath G, Markwell JP (2000) Soybean glycinin G1 acidic chain shares IgE epitopes with peanut allergen Ara h 3. *Int Arch Allergy Immunol* 123: 299-307.
 31. Ogawa T, Tsuji H, Bando N, Kitamura K, Zhu YL, et al. (1993) Identification of the soybean allergenic protein, Gly m Bd 30K, with the soybean seed 34-kDa oil-body-associated protein. *Biosci Biotechnol Biochem* 57: 1030-1033.
 32. Yaklich RW, Helm RM, Cockrell G, Herman EM (1999) Analysis of the distribution of the major soybean seed allergens in a core collection of *Glycine max* accessions. *Crop Sci* 39: 1444-1447.
 33. Tsuji H, Hiemori M, Kimoto M, Yamashita H, Kobatake R, et al. (2001) Cloning of cDNA encoding a soybean allergen, Gly m Bd 28K. *Biochim Biophys Acta* 1518: 178-182.
 34. Xiang P, Haas EJ, Zeece MG, Markwell J, Sarath G (2004) C-Terminal 23 kDa polypeptide of soybean Gly m Bd 28 K is a potential allergen. *Planta* 220: 56-63.
 35. Baur X, Pau M, Czuppon A, Fruhmant G (1996) Characterization of soybean allergens causing sensitization of occupationally exposed bakers. *Allergy* 51: 326-330.
 36. Quirce S, Fernández-Nieto M, Polo F, Sastre J (2002) Soybean trypsin inhibitor is an occupational inhalant allergen. *J Allergy Clin Immunol* 109: 178.