

Determination of Optimal Protein Quantity Required to Identify Abundant and Less Abundant Soybean Seed Proteins by 2D-PAGE and MS

Savithiry S. Natarajan · Chenping Xu · Hanhong Bae ·
Thomas J. Caperna · Wesley Garrett

Published online: 11 September 2007
© Springer-Verlag 2007

Abstract Optimizing the amounts of proteins required to separate and characterize both abundant and less abundant proteins by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is critical for conducting proteomic research. In this study, we tested five different levels of soybean seed proteins (75, 100, 125, 150, and 200 µg) by 2D-PAGE. Following 2D-PAGE and spot excision, proteins were identified by mass spectrometry analysis. The number of visible protein spots was increased with an increase in the amount of protein loaded. The intensity of highly abundant proteins [β -conglycinin β -homotrimer and glycinin G4 (A5A4B3) precursors] increased linearly between 75 and 125 µg, whereas the proglycinin G3 (A1ab1b) homotrimer showed linearity between 75 and 150 µg. The spot intensity of less abundant proteins, glycinin G2 (A2b1a) precursor and proglycinin G3 (A1ab1b) homotrimer, increased linearly with an increase in the amount of protein through 200 µg, whereas spot intensity of β -conglycinin β -homotrimer and the allergen Gly m bd 28K increased linearly until

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.

S. S. Natarajan (✉)
Soybean Genomics and Improvement Laboratory, PSI, USDA-ARS,
10300 Baltimore Avenue, Beltsville, MD 20705, USA
e-mail: savi.natarajan@ars.usda.gov

C. Xu
Department of Plant Science and Landscape Architecture, University of Maryland,
College Park, MD 20742, USA

H. Bae
Sustainable Perennial Crops Laboratory, USDA-ARS, Beltsville, MD 20705, USA

T. J. Caperna
Growth Biology Laboratory, USDA-ARS, Beltsville, MD 20705, USA

W. Garrett
Biotechnology and Germplasm Laboratory, USDA-ARS, Beltsville, MD 20705, USA

150 µg and did not increase further at 200 µg. These results suggest that 150 µg protein was a suitable amount for the separation of abundant proteins, and 200 µg protein was suitable for the separation of less abundant proteins prepared from soybean seeds.

Keywords Proteomics · Soybean seeds · *Glycine max* · MALDI-TOF-MS · LC-MS/MS · 2D-PAGE · Glycinin · β-conglycinin

Introduction

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and mass spectrometry (MS)-based proteomic technologies are powerful tools for the analysis of complex protein mixtures. These techniques have been successfully used for genetic and proteomics studies of plant materials (Lottspeich 1999; Gorg et al. 2000). 2D-PAGE systems were used to separate various globulin proteins from soybean seed (Mei-Guey et al. 1983; Herman et al. 2003; Natarajan et al. 2006a). However, determination of protein extraction techniques and optimization of protein quantities required to distinctly separate both abundant and less abundant proteins by 2D-PAGE remain as challenges to the use of this technology for bean seeds because of the presence of a large proportion of a few very abundant proteins. In soybean, storage proteins account for about 70–80% of the total seed protein. A number of recent advances in 2D-PAGE methodologies allow more proteins to be arrayed in micropreparative quantities (Herbert 1999; Gorg et al. 2000). To evaluate the maximum number of proteins when comparing genetically distinct soybean variants, we have determined the optimal quantity of proteins that allows analysis of abundant and less abundant proteins. In this type of analysis, it is not only important to have good separation, but it is necessary to determine the relative intensity of protein spot when both abundant and less abundant proteins are examined on the same gel. Among several proteomic tools, matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) and liquid chromatography mass spectrometry (LC-MS/MS) are sensitive methods for accurately characterizing and identifying proteins separated by 2D-PAGE (Karas and Hillenkamp 1988; Hunt et al. 1986). In this study, we compared five different quantities of soybean seed proteins using 2D-PAGE analysis, and identified the proteins using MS analysis.

Materials and Methods

Plant Materials Soybean seeds [*Glycine max* (L.) Merr.] of cultivar Williams 82 were obtained from the USDA soybean germplasm collection, Urbana, IL. Seeds were stored at –80°C until used.

Extraction of protein

The extraction of protein was performed according to Natarajan et al. (2005).

- Powder the soybean seeds in liquid nitrogen using mortar and pestle.
- Homogenize 100 mg of the soybean seed powder with 5 ml of a solution containing 10% (w/v) trichloroacetic acid (TCA) in acetone with 0.07% (v/v) 2-mercaptoethanol.

- Precipitate the total protein for 1 h at -20°C .
- Centrifuge the extract at $20,800\times g$ for 20 min at 4°C .
- Wash the pellet two to three times with acetone containing 0.07% (v/v) 2-mercaptoethanol and dry under vacuum for 30 min.
- Resuspend the pellet in 1 ml of lysis buffer [9 M urea, 1% CHAPS, 1% (w/v) ampholytes (pH 3–10), and 1% dithiothreitol (DTT)] followed by sonication for 15 min.
- Remove the insoluble material by centrifugation at $20,800\times g$ for 20 min at 4°C and use the supernatant for 2D-PAGE analysis.

Protein Determination and Electrophoresis

- Determine the amount of proteins with the Bradford method (Bradford 1976) using a commercial dye reagent (Bio-Rad, Hercules, CA, USA). Precipitate the sample in 10% (w/v) TCA, centrifuge, and resolubilize protein pellet in 1 N NaOH prior to analysis. Determine the protein content based on bovine serum albumin standard.
- Perform the first-dimension isoelectric focusing (IEF) using 13 cm pH 3–10 linear immobilized pH gradient (IPG) strips in the IPGphor system (GE Healthcare, Piscataway, NJ, USA). Rehydrate the IPG strips with 250 μl rehydration buffer (8 M urea, 2% CHAPS, 0.5% ampholytes, 0.002% bromophenol blue) containing the appropriate amount of protein.
- Voltage settings for isoelectric focusing are 500 V for 1 h, 1,000 V for 1 h, and 8,000 V, to a total 14.5 kVh.
- Electrophorese the focused strips either immediately on a second-dimension gel or store at -80°C for future use.
- For the second-dimension gel electrophoresis, incubate the gel strips with equilibration buffer 1 [50 mM Tris-HCl pH 8.8, 6 M urea, 30% glycerol, 2% sodium dodecyl sulfate (SDS), 0.002% bromophenol blue, 1% DTT] and equilibration buffer 2 (50 mM Tris-HCl pH 8.8, 6 M urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue, 2.5% iodoacetamide) for 15 min each.
- Place the strips onto 12% polyacrylamide gels (18 $\times$ 16 cm) with Tris-glycine buffer system as described by Laemmli (1970).
- Overlay the strips with agarose sealing solution (0.25 M Tris base, 1.92 M glycine, 1% SDS, 0.5% agarose, and 0.002% bromophenol blue).
- Perform the electrophoresis using the Hoefer SE 600 Ruby electrophoresis unit (GE Healthcare) according to the manufacturer's recommendations.
- Fix the gels overnight in 50% ethanol and 10% acetic acid and wash 3 $\times$ 30 min with distilled water.
- Visualize the 2D-PAGE gels by staining with Colloidal Coomassie Blue G-250 as described by Newsholme et al. (2000). Pretreat the gels for 1 h in 34% methanol, 17% ammonium sulfate, and 3% phosphoric acid. Finally, stain the gels in the same solution containing Coomassie Blue G-250 (0.066%) for 2 days.
- Store the gels in 20% ammonium sulfate solution and scan using laser densitometry (PDSI, GE Healthcare).
- Use triplicate samples for soybean seed protein extraction and 2D-PAGE analysis.

Image Analysis of Protein Spots

- Perform image analysis with ImageQuant™ for Windows™ NT software (GE Healthcare), which utilizes area and profile-based tools.
- Convert the gel image file to tagged image file format (.tif) and use the software drawing tools to encircle detectable spots to specify regions for analysis and to subtract background signal.
- Automatic algorithms calculate the quantitative statistics, including spot volume, pixel area, mean pixel intensity, standard deviation, and background values.

In-Gel Digestion of Protein Spots

To identify protein spots by MS analysis, excise each spot manually from the stained gel.

- Wash the gel first with distilled water to remove ammonium sulfate and then with 50% acetonitrile containing 25 mM ammonium bicarbonate to destain the gel plug.
- Dehydrate the gel plug with 100% acetonitrile, dry under vacuum, and then rehydrate with 20 μ l of 10 μ g/ml trypsin (modified porcine trypsin, sequencing grade, Promega, Madison, WI, USA) in 25 mM ammonium bicarbonate.
- Perform the digestion overnight at 37°C.
- Extract the resulting tryptic fragments with 50% acetonitrile and 5% trifluoroacetic acid using sonication.
- Dry the extract to completeness and then dissolve in 50% acetonitrile and 0.1% trifluoroacetic acid.

Protein Identification

- Prepare protein samples for MALDI-TOF-MS (Applied Biosystems, Framingham, MA, USA) using the dried droplet method (Karas and Hillenkamp 1988) with an α -cyanohydroxycinnamic acid matrix.
- Acquire the spectra with 50 shots of a 337-nm nitrogen laser operating at 20 Hz.
- Calibrate the spectra using the trypsin autolysis peaks at mass-to-charge ratio (m/z) 842.51 and 2,211.10 as internal standards.
- For LC-MS/MS, separate the peptides on a reverse-phase column using a 30-min gradient of 5 to 60% acetonitrile in water with 0.1% formic acid.
- Operate the instrument with a duty cycle that acquires MS/MS spectra on the three most abundant ions identified by a survey scan from 300 to 2,000 Da.
- Apply dynamic exclusion to prevent the continuous analysis of the same ions.
- Place the parent mass on an exclusion list for the duration of 1.5 min and process the raw data by Sequest to generate DTA files for database searching.
- Use the merge.pl script from Matrix Science to convert multiple Sequest DTA files into a single Mascot generic file suitable for searching in Mascot.

Data Interpretation Using Databases

- Determine peptide fragments of the trypsin digested protein spots by MALDI-TOF-MS.

- Use theoretically generated peptide masses to determine m/z values from known protein sequences (Yates 1993; Eng et al. 1994).
- Perform protein identification by searching the National Center for Biotechnology Information (NCBI) nonredundant database using the Mascot search engine (<http://www.matrixscience.com>), which uses a probability-based scoring system (Perkins et al. 1999).

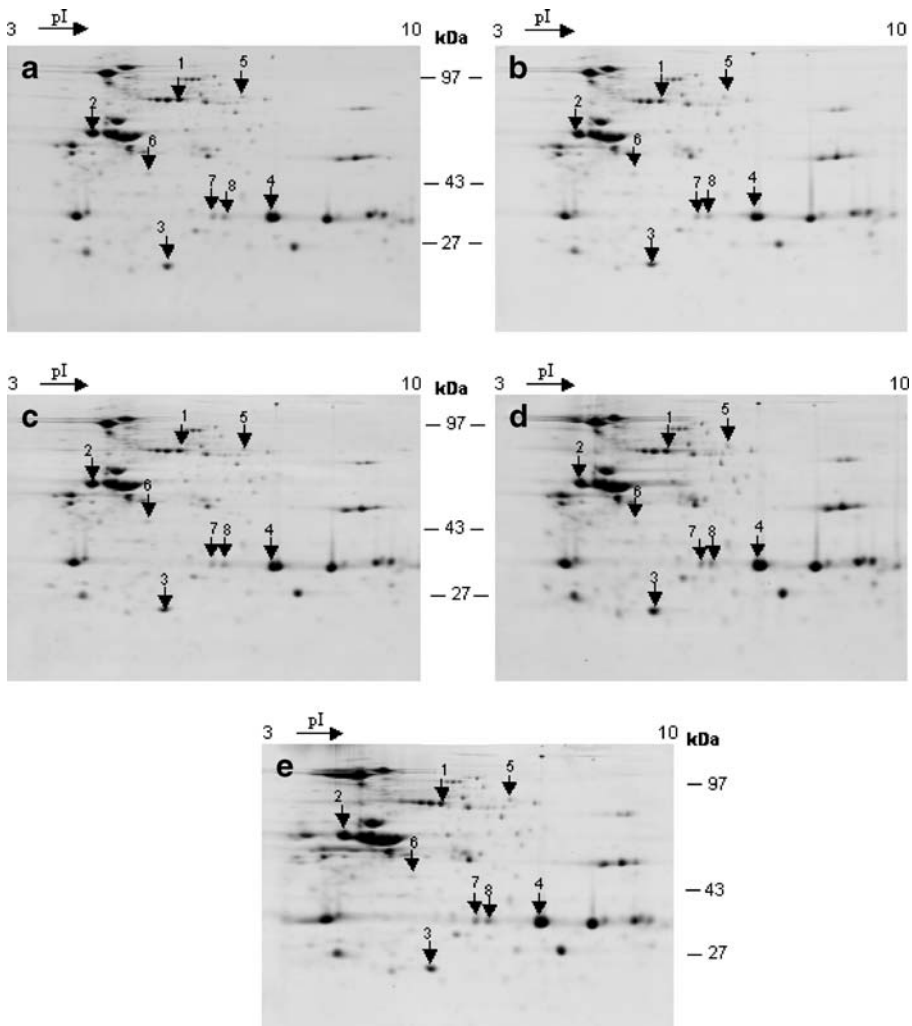


Fig. 1 Comparison of different protein loading quantities on 2D-PAGE separation of abundant and less abundant soybean proteins extracted using a modified trichloroacetic acid (TCA)–acetone precipitation–urea solubilization extraction buffer. **a** 75 µg of protein; **b** 100 µg of protein; **c** 125 µg of protein; **d** 150 µg of protein; **e** 200 µg of protein. The first-dimension run using pH 3–10 immobilized pH gradient (IPG) strips and the second dimension is a 12% SDS-PAGE. Gels were stained with colloidal Coomassie Blue stain G-250. Arrows indicate protein spots analyzed in this investigation (see Table 1)

Results and Discussion

In this study, we compared five different protein amounts, 75, 100, 125, 150, and 200 μg , of soybean seed proteins to separate and analyze both abundant and less abundant proteins by 2D-PAGE. Among these five amounts (Fig. 1a–e), we observed that 100 μg (Fig. 1b: Fig. 2a) of protein was a suitable amount for separating a large number of highly abundant proteins. Similarly, 200 μg of protein (Fig. 1e: Fig. 2b) was suitable for the separation of a large number of less abundant proteins. Herman et al. (2003) used 120 μg of proteins and separated soybean seed proteins; however, no attempt was made to determine relative spot intensities. The total number of protein spots, analyzed and counted using Image Quant, increased as the amount of protein was increased from 75 to 200 μg . Specifically, totals of 174, 200, 224, 251, and 282 protein spots were found on the 75, 100, 125, 150, and 200 μg protein gels, respectively ($P < 0.001$, $r^2 = 0.99$).

To identify specific abundant and less abundant proteins following 2D-PAGE, spots were manually picked from Coomassie-stained gels and analyzed by MS. We previously identified 2D-PAGE separated soybean seed proteins from different classes, storage, allergen, and antinutritional by MS (Natarajan et al. 2006a, b, 2007). In this study, we have compared the intensity of four representative abundant

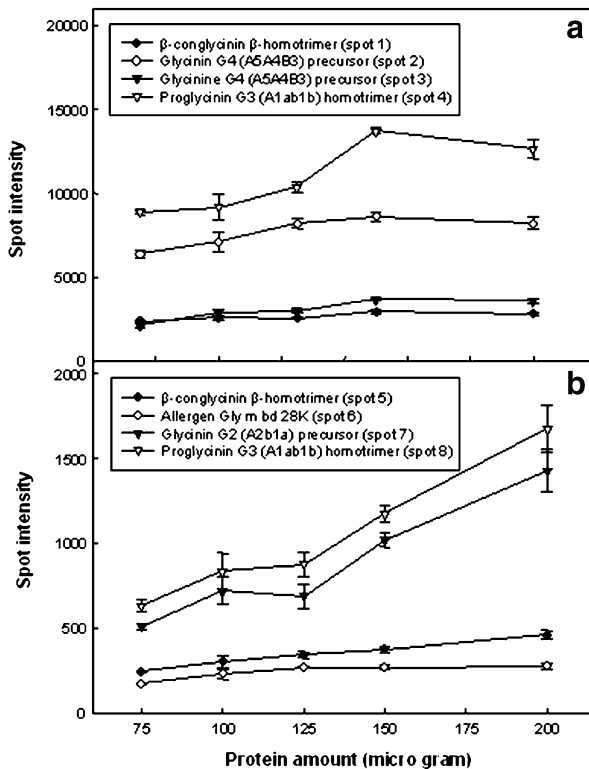


Fig. 2 **a** Intensity of abundant protein spots at different protein quantities. **b** Intensity of less-abundant protein spots at different protein quantities

protein spots (# 1, 2, 3, and 4; Table 1) and four representative less abundant protein spots (# 5, 6, 7 and 8; Table 1). Table 1 contains the assigned protein spot number, calculated isoelectric point, molecular weight, number of peptides matched, percent sequence coverage, MOWSE score, expect value, and NCBI database accession number of the best match for each of these eight spots.

The intensity of the abundant protein spots, spot #1 (β -conglycinin beta-homotrimer) and spots #2 and #3 [glycinin G4 (A5A4B3) precursor], increased linearly between 75 and 125 μ g of protein quantities, although at higher amounts, the intensity did not increase proportionately (Fig. 2a). The highly abundant spot #4 [proglycinin G3 (A1ab1b) homotrimer] appeared to be most intense at 150 μ g, and thereafter sharply declined. This apparent decline could be due to the threshold limit of the ImageQuant. Except for spot #6 (allergen Gly m Bd 28k), the intensities of four less abundant proteins, spot #5 (β -conglycinin beta-homotrimer), spot #7 [Glycinin G2 (A2B1a) precursor], and spot #8 [proglycinin G3 (A1ab1B) homotrimer], increased linearly with an increase in the amount of protein separated throughout the range of protein quantities tested (Fig. 2b; Table 1). However, the spot intensity of glycinin G2 (A2B1a) precursor, spot #7, and the proglycinin G3 (A1ab1B) homotrimer, spot #8, did not increase linearly at the lower quantities tested (Fig. 2b). The intensity of the allergen Gly m Bd 28k, spot # 6, increased linearly until 125 μ g, with no change thereafter.

This investigation demonstrates that 100 to 150 μ g protein is better for the separation of highly abundant proteins in terms of spot intensity as well as total number of spots, whereas between 150 and 200 μ g protein was suitable for the analysis of less abundant proteins. The identities of these proteins by their approximate molecular weights and isoelectric points were similar to those of previous reports (Schuler et al. 1982; Herman et al. 2003; Mooney and Thelen 2004; Natarajan et al. 2005, 2006a, 2007), confirming the reproducibility of this 2D-PAGE protocol.

Table 1 Proteins identified by MALDI-TOF-MS and LC-MS analysis

Spot ID	Calculated PI/Mr	Protein identity	Peptides matched	Sequence coverage (%)	MOWSE score	Expect	NCBI accession number
1	5.67/47879	β -conglycinin β -homotrimer	26	49	240	1.50E-19	gi 21465628
2	5.38/64097	Glycinin G4 (A5A4B3) precursor	13	16	76	4.20E-03	gi 99910
3	5.38/64136	Glycinin G4 (A5A4B3) precursor	6	8	160		gi 99910
4	5.78/54047	Proglycinin G3 (A1ab1b) homotrimer	8	18	72	9.40E-03	gi 15988117
5	5.67/47879	β -conglycinin β -homotrimer	21	45	132	9.70E-09	gi 21465628
6	5.73/52813	Allergen Gly m Bd 28 K	3	6	187		gi 12697782
7	5.56/54903	Glycinin G2 (A2b1a) precursor	9	14	313		gi 72295
8	5.78/54047	Proglycinin G3 (A1ab1b) homotrimer	9	18	116	3.90E-07	gi 15988117

Acknowledgements Dr. J. Slovin and Dr. D. Lakshman are thanked for their critical review of this manuscript. Funding for this research was provided by ARS project 1275-21000-223-00D

References

- Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal Biochem* 1976;72:248–54.
- Eng J, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein databases. *J Am Soc Mass Spectrom* 1994;5:976–89.
- Gorg C, Obermaier G, Boguth A, Harder B, Scheibe R, Wildgruber, et al. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 2000;21:1037–53.
- Herbert B. Advances in protein solubilization for two-dimensional electrophoresis. *Electrophoresis* 1999;20:660–3.
- Herman EM, Helm RM, Jung R, Kinney AJ. Genetic modification removes an immunodominant allergen from soybean. *Plant Physiol* 2003;132:36–43.
- Hunt DF, Yates JR III, Shabanowitz J, Winston S, Hauer CR. Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A* 1986;83:6233–7.
- Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10000 Dalton. *Anal Chem* 1988;60:2299–301.
- Laemmli U. Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature* 1970;227:680–5.
- Lottspeich F. Proteome analysis: a pathway to the functional analysis of proteins. *Angew Chem Int Ed Engl* 1999;38:2476–92.
- Mei-Guey L, Tyrell R, Bassette R, Reeck GR. Two dimensional electrophoretic analysis of soybean proteins. *J Agric Food Chem* 1983;31:963–8.
- Mooney BP, Thelen JJ. High-throughput peptide mass fingerprinting of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. *Phytochemistry* 2004;65:1733–44.
- Natarajan SS, 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.
- 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 proteomics analysis. *J Agric Food Chem* 2006a;54:3114–20.
- Natarajan SS, Xu C, Bae H, Caperna TJ, Garrett WM. Proteomic analysis of allergen and anti-nutritional proteins in wild and cultivated soybean. *J Plant Biochem Biotech* 2006b;15:103–8.
- Natarajan SS, Xu C, Bae H, Bailly BA, Cregan P, Caperna TJ, et al. Proteomic and genetic analysis of glycinin, subunits of sixteen soybean genotypes. *Plant Physiol Biochem* 2007;45:436–44.
- Newsholme SJ, Malefyt BF, Steiner S, Anderson NL, Schwartz LW. Two-dimensional electrophoresis of liver proteins: Characterization of a drug-induced hepatomegaly in rats. *Electrophoresis* 2000;21:2122–8.
- Perkins DN, Pappinl DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999;20:3551–67.
- Schuler MA, Ladin BF, Fryer G, Pollaco JG, Beachy RN. Structural sequences are conserved in the genes coding for the α' , α and β subunits of the soybean seed storage protein and other seed proteins. *Nucleic Acids Res* 1982;10:8245–60.
- Yates JR. Mass spectrometry and the age of the proteome. *J Mass Spectrom* 1993;33:1–19.