CHAPTER TWELVE

Soybean seed proteomics: Methods for the isolation, detection, and identification of low abundance proteins

Wonseok Kim[a,†], Joel Nott[b,†], Sunhyung Kim[a], and Hari B. Krishnan[a,c,*]
[a]Division of Plant Science and Technology, University of Missouri, Columbia, MO, United States
[b]The Protein Facility of the Iowa State University Office of Biotechnology, Ames, IA, United States
[c]Plant Genetics Research Unit, USDA, Agricultural Research Service, Columbia, MO, United States
*Corresponding author: e-mail address: hari.b.krishnan@usda.gov

Contents

Abstract

The salt-soluble globulins, glycinins (11S globulin), and β-conglycinins (7S globulin), are the most abundant seed proteins of soybean seeds. Together, these two groups of proteins account for 60–70% of total soybean seed proteins. Proteomic assessment of the less abundant soybean seed proteins using general isolation protocols is

† These authors contributed equally.

Methods in Enzymology, Volume 676
ISSN 0076-6879
https://doi.org/10.1016/bs.mie.2022.07.001

challenging due to the overwhelming abundance of storage proteins. Development of a simple, fast, and inexpensive method to remove most storage proteins from a seed extract will significantly enhance the study of the nonabundant proteins within seeds. We have developed two simple methods for the depletion of abundant seed proteins resulting in the enrichment of low abundance proteins from soybean seeds. Here, we provide a detailed procedure for the isolation, separation, identification, and quantification of low abundance seed proteins of soybean.

1. Introduction

Seeds have played a vital role in human diet as a major source of dietary protein. For this reason, seeds of crop plants have been intensively studied. Seeds store large amount of reserve materials in form of proteins, carbohydrates, and lipids that are mobilized during seed germination. Among crop plants, legumes have been extensively used as a major source of protein because they accumulate higher concentration of proteins than cereals. Seeds accumulate various proteins, most of which have either metabolic or structural roles. In addition, the seeds commonly accumulate a group of abundant proteins termed as storage proteins (Shewry, Napier, & Tatham, 1995), which are responsible for determining overall total protein content of the seed. Because of their abundance and economic importance, seed storage proteins were among the earliest proteins that were isolated and crystalized (Shewry et al., 1995). The systematic study of seed proteins was first initiated by Osborne (1924), who classified them into four groups (albumins, globulins, prolamins, and glutelins) based on their solubility properties. From that point onwards, numerous analytical, biochemical, and physiological methods have been extensively used to characterize seed proteins from a wide variety of source plants. The advent of "omics" technology has immensely accelerated the identification and characterization of seed storage proteins.

Proteomics is the large-scale study of proteomes. Seed proteomics has progressed rapidly due to improvements in instrumentation and refinements

in protein isolation and separation techniques and technologies (Miernyk, 2014; Miernyk & Hajduch, 2011; Smolikova et al., 2020). The proteome of seeds, which reflects the underlying transcriptome, differs from each other. For example, the seed proteome of maize is quite distinct from that of pea (Bourgeois et al., 2009; Yu et al., 2016). Proteomics is often used for investigating the rates of protein production, steady-state abundance, post-translational modifications, protein-protein interactions, and to identify proteins involved in specific biological process (Low et al., 2021; McWhite et al., 2020; Zecha et al., 2022). Progress in protein separation techniques, in conjunction with technological advances in mass spectrometry, has allowed proteomics to develop quickly, where it is now playing an increasingly key role in functional genomics (Smolikova et al., 2020; Zhang et al., 2022).

Soybean (*Glycine max* [L.] Merr) seeds feature exceptionally high levels of protein (~40%) and is the primary source of protein for animal feed worldwide (Cabanos, Matsuoka, & Maruyama, 2021; Krishnan, 2000; Nielsen, 1996). Soybean seed proteins are composed of two salt-soluble protein classes: glycinins (11S globulin) and β-conglycinins (7S globulin) (Derbyshire, Wright, & Boulter, 1976). Both the glycinins and β-conglycinins are encoded by multigene families (Cho, Davies, & Nielsen, 1989; Li & Zhang, 2011). Six functional genes encoding glycinin (Gy1, Gy2, Gy3, Gy4, Gy5, and Gy7) have been reported in soybean (Beilinson et al., 2002; Nielsen et al., 1989). Each glycinin is synthesized as a precursor protein, which is post-translationally cleaved into acidic and basic polypeptides that are linked by a disulfide bond (Staswick, Hermodson, & Nielsen, 1984). In contrast to other glycinins, the glycinin produced by Gy4 undergoes cleavage at two sites resulting in two acidic (A4 and A5) peptides and one basic (B3) peptide (Momma et al., 1985) with theoretical molecular masses of 11 kDa (A5), 29 kDa (A4), and 21 kDa (B3), respectively (Momma et al., 1985). As many as 15 β-conglycinin genes have been identified in soybean cultivar "Williams 82" genome (Li & Zhang, 2011); however, most of these genes appear to be pseudogenes and only three genes encode the most highly seed-expressed α′-, α-, and β-conglycinins (Kitamura, Takagi, & Shibasaki, 1976; Thanh & Shibasaki, 1978).

Together, the 7S and 11S globulins account for 60–70% of soybean total seed proteins (Krishnan, 2000; Nielsen, 1996). The other seed proteins only account for a minor portion of the total seed protein content. Recent advances in mass spectrometry have led to the identification of numerous soybean seed proteins that have significantly increased the proteome coverage of soybean seed (Agrawal, Hajduch, Graham, & Thelen, 2008; Hajduch, Ganapathy, Stein, & Thelen, 2005; Islam et al., 2019; Islam, Krishnan, & Natarajan, 2020, 2021; Komatsu et al., 2017; Min, Hyeon, et al., 2020; Mooney, Krishnan, & Thelen, 2004; Ohyanagi, Sakata, & Komatsu, 2012). However, analysis of soybean proteome consisting of thousands of proteins has limited dynamic resolution because only the most abundant proteins can be detected (van Wijk, 2001).

Proteomic assessment of the less abundant proteins within soybean seed using general isolation protocols is challenging when the overwhelming majority, sometimes 60–80%, is made up of storage proteins. Development of a simple, fast, and inexpensive method to remove most storage proteins from a seed extract would significantly enhance the study of the nonabundant proteins within seeds. Methods have been reported in the literature for the depletion of abundant seed proteins based upon different physiological and/or biochemical principles (Kim et al., 2015; Miernyk & Hajduch, 2011; Miernyk & Johnston, 2006; Min, Park, et al., 2020). Our laboratory has developed simple, rapid, and inexpensive methods to isolate and enrich low abundant proteins from soybean seeds (Krishnan, Oehrle, & Natarajan, 2009; Natarajan, Krishnan, Lakshman, & Garrett, 2009). Here, we provide a detailed protocol for the isolation, separation, identification, and quantification of low abundance seed proteins of soybean. A schematic outline for the isolation, separation, and proteomic analysis of low abundant soybean seed proteins is shown in Fig. 1.

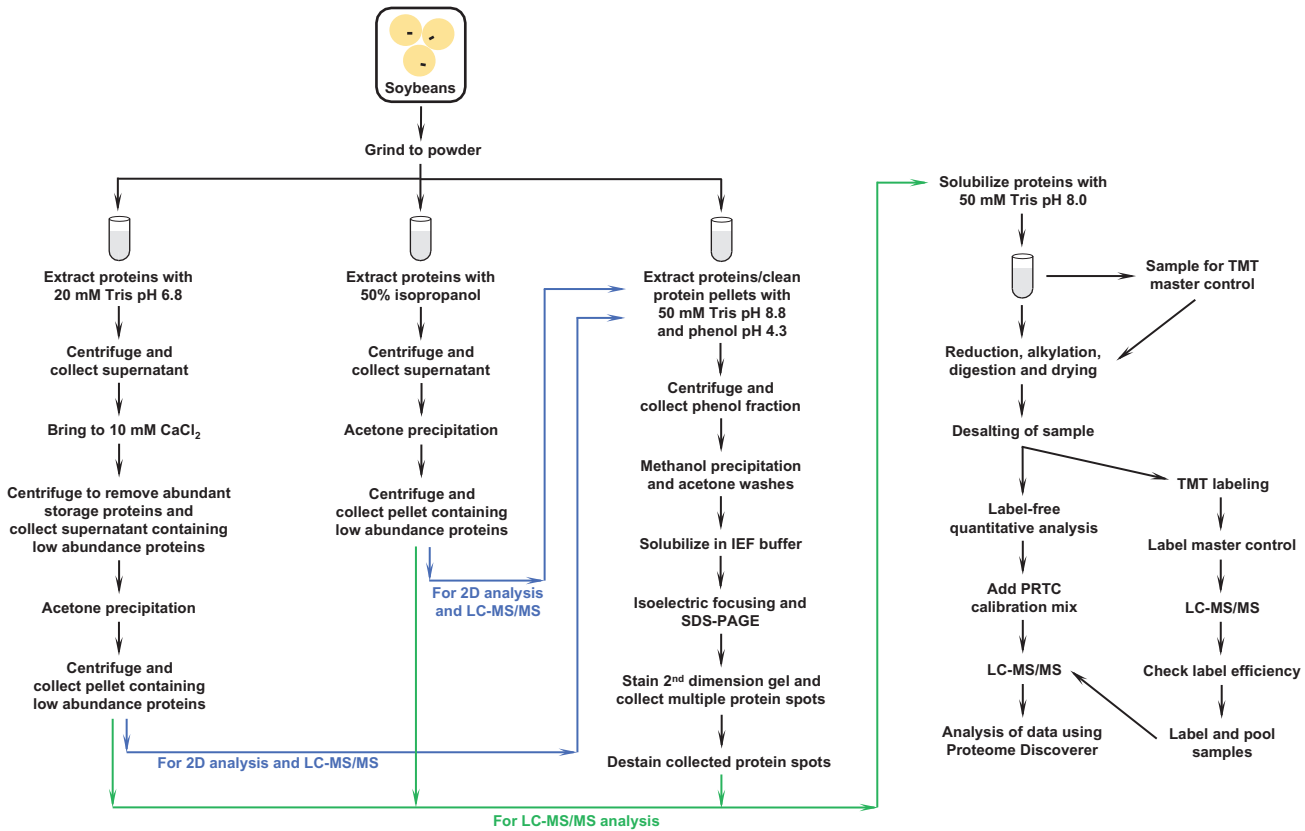

Fig. 1 A schematic outline for the isolation, separation, and proteomic analysis of low abundance soybean seed proteins.

2. $CaCl_2$ fractionation method to obtain low abundance proteins

Depletion of abundant seed storage proteins has been achieved, with different degree of success, through various strategies based on immunoremoval, isoelectric precipitation, salting out, affinity chromatography, random peptide-beads, and fractional solubility (Kim et al., 2015; Miernyk & Hajduch, 2011; Miernyk & Johnston, 2006). Inherent difficulties are encountered with most of these techniques due to low efficiency, cross-reactivity, broad-range affinity, low reproducibility, and expense of technique. Among these methods, the calcium chloride ($CaCl_2$) fractionation protocol developed in our laboratory to selectively remove storage proteins is the easiest, has the greatest capacity and efficiency, and is the least expensive. Our method is quick and can be performed with common laboratory equipment and reagents.

2.1 Materials and equipment

Tris base [Tris(hydroxymethyl)aminomethane]
Centrifuge capable of achieving 16,000 x *g*
Centrifuge tubes capable of withstanding 16,000 x *g*
100 mM $CaCl_2 \cdot 2H_2O$ (calcium chloride dihydrate)
100% Acetone
−20 °C freezer

2.2 Protocol

1. Grind soybean seeds into a fine powder using a mortar and pestle
2. Weigh out 200 mg of seed powder and transfer into a 15 mL centrifuge tube
3. Add 10 mL of 20 mM Tris-HCl, pH 6.8
4. Mix the tube contents for 10 min, with vigorous shaking, at room temperature
5. Clarify the slurry by centrifugation at 16,000 × *g* for 15 min
6. Remove the supernatant and place it into a clean tube
7. Add the required amount of 100 mM $CaCl_2$ stock solution to obtain a final concentration of 10 mM
8. Mix the tube contents for 1 min and leave it at room temperature for 5 min
9. Clarify the slurry by centrifugation at 16,000 × *g* for 15 min

10. Discard the pellet that contains predominantly the abundant seed storage proteins
11. Carefully remove the supernatant, which contains less abundant seed proteins, and place it into a clean tube
12. To this, add 3 volumes of ice-cold 100% acetone, mix the contents, and place it at −20 °C freezer overnight
13. Centrifuge the samples at 16,000 × *g* for 15 min
14. Air-dry the pellet at room temperature for 10 min
15. For one-dimensional (1D) electrophoretic gel analysis (Fig. 2), the pellet can be directly dissolved in SDS-sample buffer. For two-dimensional (2D) electrophoretic analysis gel analysis (Fig. 3), re-isolate the pellet from Step 14 beginning with Step 2 (Section 4.1.2).

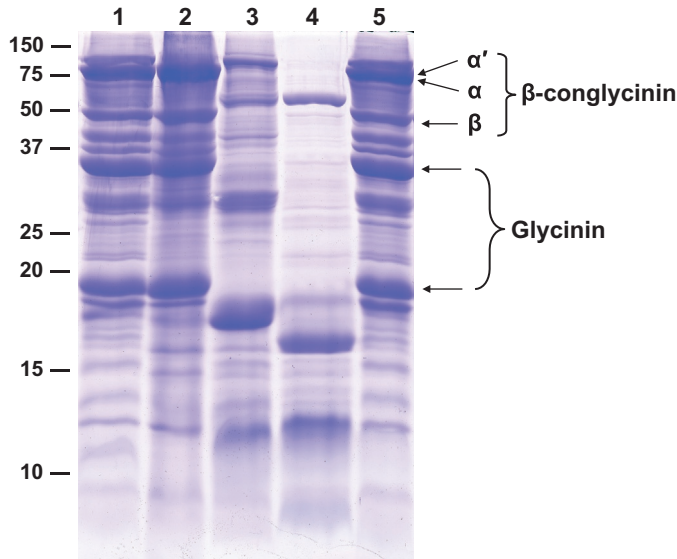

Fig. 2 One-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (1D SDS-PAGE) analysis of soybean seed proteins. Total soybean seed proteins (lanes 1 and 5), soybean storage proteins specifically removed by calcium precipitation (lane 2), low abundance soybean seed proteins obtained by calcium fractionation (lane 3), and 50% isopropanol extracted soybean seed proteins (lane 4) were analyzed by 15% SDS-PAGE. Separated proteins were visualized by staining the gel with Coomassie Blue. Polypeptides corresponding to α′, α and β-subunits of β-conglycinin and the acidic and basic subunits of glycinin are shown with arrows. The position and sizes of protein markers in kDa are shown on the left side of the figure.

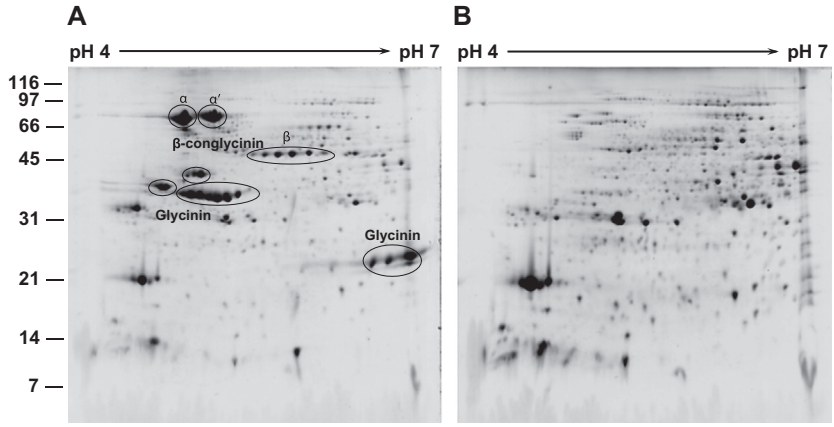

Fig. 3 Two-dimensional (2D) gel electrophoresis separation of soybean seed proteins. Total (A) and calcium chloride fractionated (B) soybean seed proteins were separated by isoelectric focusing on IPG strips (pH 4–7) and then by SDS-PAGE on 13.5% gels. The gels were stained with Colloidal Coomassie Blue G-250. Calcium chloride fractionation removes approximately 90% of highly abundant seed proteins from the protein extract (A), allowing for nearly 300 previously inconspicuous proteins present in soybean seed to be more detectable (B). Protein spots corresponding to α′, α and β-subunits of β-conglycinin and the acidic and basic subunits of glycinin are circled (A). The position and sizes of protein markers in kDa are shown on the left side of the figure.

16. For mass spectrometry analysis, resuspend the pellet from Step 14 in minimal volume of 50 mM Tris-HCl, pH 8.0, and proceed with Step 1 (Section 4.2.2).

3. Isopropanol extraction method to obtain low abundance proteins

In addition to the calcium chloride fractionation protocol in Section 2, we developed another method in our laboratory to selectively remove storage proteins. Isopropanol fractionation is also easy, has great capacity and efficiency, and is inexpensive. The protein pellet obtained from the final protocol step can be further processed for either gel electrophoresis or mass spectrometry.

3.1 Materials and equipment

50% (*v*/v) Isopropanol

100% Acetone

Centrifuge capable of achieving 16,000 × *g*

Centrifuge tubes capable of withstanding 16,000 × *g*
−20 °C freezer

3.2 Protocol

1. Grind soybean seeds into a fine powder using a mortar and pestle
2. Weigh out 500 mg of seed powder and transfer into a 15 mL tube containing 10 mL of 50% isopropanol
3. Mix the tube contents for 15 min, with vigorous shaking, at room temperature
4. Clarify the slurry by centrifugation at 16,000 × *g* for 15 min
5. Remove the clear supernatant, measure the volume, and place it into a clean tube
6. To this, add 3 volumes of ice-cold acetone, mix the contents, and place it at −20 °C freezer overnight
7. Centrifuge the samples at 16,000 × *g* for 15 min
8. Air-dry the pellet at room temperature for 10 min
9. For 2D gel analysis, re-isolate the pellet from Step 8 beginning with Step 2 (Section 4.1.2).
10. For mass spectrometry analysis, resuspend the pellet from step 8 in minimal volume of 50 mM Tris-HCl, pH 8.0, and proceed with Step 1 (Section 4.2.2).

4. Proteomic analysis of low abundance soybean seed proteins

4.1 Two-dimensional gel electrophoresis

For 2D electrophoretic gel analysis, fractionated proteins were separated using the method outlined below. Protein concentration estimate, after solubilized in isoelectric focusing (IEF) resuspension buffer was performed using the Bradford method, and 300 µg of protein sample loaded per strip using overnight in-gel rehydration.

4.1.1 Materials and Equipment

Tris base [Tris(hydroxymethyl)aminomethane]
Glycine
Acrylamide/Bis-acrylamide
Ammonium persulfate
N,N,N′,N′-Tetramethylethylenediamine (TEMED)
Coomassie Brilliant Blue G-250

Urea
Thiourea
Methanol
Ammonium acetate
β-mercaptoethanol
Glycerol
Agarose (low EEO)
Saturated phenol (pH 4.3)
3-[(3-Cholamidopropyl)-dimethylammonio]-1-propane-sulfonate (CHAPS)
Sucrose
Acetic acid
3-(4-Heptyl)phenyl-3-hydroxypropyl)dimethyl-ammoniopropanesulfonate) (C7BzO)
2-hydroxy-ethyldisulfide (2-HED)
Dithiothreitol (DTT)
Iodoacetamide (IAM)
Acetonitrile (ACN)
Ammonium bicarbonate
Protease inhibitor cocktail (Plant ProteaseArrest; G-Biosciences, St. Louis, MO, USA)
Immobilized pH Gradient (IPG) strips (Cytiva, Marlborough, MA, USA)
Centrifuge capable of achieving $14{,}000 \times g$
Swing-bucket centrifuge capable of achieving $5000 \times g$
Centrifuge tubes capable of withstanding $14{,}000 \times g$ and resistant to phenol
−20 °C freezer and − 80 °C freezer

4.1.2 Protocol

1. For total protein extraction from soybean, grind seed to a fine powder using a mortar and pestle
2. To this powder, or to the protein pellet from prior fractionation, add 5 mL of 0.1 M Tris-Cl, pH 8.8, containing 0.9 M sucrose, 0.2 M NaCl, 0.4% (*w*/*v*) β-mercaptoethanol, and protease inhibitor cocktail, and continue grinding
3. Add 5 mL of saturated phenol, pH 4.3, and continuously mix for 30 min at room temp. *Note: The volume ratio of 1:1 (buffer:phenol) is critical*

4. Obtain the phenolic phase by centrifugation at 5000 × *g* for 20 min at 23 °C in a swing-bucket rotor
5. The upper phenolic phase was removed and added to ten volumes of 100% methanol with 0.1 M ammonium acetate. Proteins were allowed to precipitate for 2 h at −80 °C
6. Centrifuge at 14,000 × *g* for 15 min at 4 °C. Wash this protein pellet twice more in a solution of 100% methanol, 0.1 M ammonium acetate, with 0.01 M DTT. Repeat resuspension/chilling/centrifugation of the insoluble proteins 3 × additionally using a solution of 100% acetone, with 0.01 M DTT
7. Following the final wash, the protein pellet was allowed to air dry slightly, and then dissolved in a small volume of IEF resuspension buffer (7 M urea, 2 M thiourea, 1% (*w/v*) CHAPS, and 2% (w/v) C7BzO). Obtain the protein concentration using Bradford method
8. Rehydrate IPG strips with 250 μL of the IEF resuspension buffer, also containing 300 μg of your solubilized protein, along with a final concentration of the following: 5% (*v*/v) glycerol, 2.2% (v/v) 2-hydroxyethyl disulfide, 60 mM DTT, and 0.3% (v/v) IPG buffer (carrier ampholytes).
9. Rehydrate the strips at 22 °C for 15 h prior to focusing
10. Conduct isoelectric focusing as follows: 50 V active rehydration, 2 h; 250 V, 500 Vhrs, fast ramp; 1000 V, 500 Vhrs, fast ramp; 8000 V, 2 h, linear ramp; 8000 V, 60,000–80,000 Vhrs, fast ramp. *Note: Volt hours (Vhrs) depends on pH gradient*
11. Prior to the second dimension, equilibrate the IPG strips with 5% SDS in a urea-based solution (50 mM Tris-Cl pH 8.8, 6 M urea, 30% (*v*/v) glycerol and 0.1% (*w/v*) bromophenol blue) containing 2% (*w/v*) DTT for 15 min
12. Following step 10, equilibrate the IPG strips with 5% SDS in a urea-based solution (50 mM Tris-Cl pH 8.8, 6 M urea, 30% (v/v) glycerol and 0.1% (*w/v*) bromophenol blue) containing 2.5% (w/v) IAM for 15 min
13. Carefully place the IPG strips onto a 15% (w/v) Tris-glycine vertical second dimension and secure into place with warm 1% (w/v) agarose dissolved in SDS-PAGE running buffer (containing 0.2% (w/v) SDS).
14. Run gels at an initial 10 mA/gel for 1 h followed by 25 mA/gel for the remainder of the run (elimination of dye front; approximately 4 h).
15. Remove gels from the cassette and soak them in solution of 5:4:1 (methanol:water:acetic acid) for 30 min

16. Remove the gels from this solution, rinse twice with ultrapure water for 10 min each wash, and stain them with Coomassie G-250 for overnight
17. Destain the gels with ultrapure H_2O for several hours and scan the gel
18. Excise a small gel piece of each protein spot using a 1–1.5 mm spot picker
19. Wash gel plugs briefly in distilled water and then destain completely in a 50% (*v*/v) solution of acetonitrile/water containing 25 mM ammonium bicarbonate. After a 100% acetonitrile wash, the protein contained in the acrylamide gel can be stored at −80 °C freezer
20. Proceed to Section 4.2.1 for mass spectral analysis

4.2 Mass spectral analysis

All solvents should be LC-MS grade and only ultrapure HPLC grade water should be used.

4.2.1 Materials and equipment

SpeedVac Concentrator (Savant)
Thermomixer C with ThermoTop (Eppendorf)
Acetonitrile (ACN)
Formic acid (FA)
Ultrapure water

4.2.1.1 Digestion

Tris base [Tris(hydroxymethyl)aminomethane]
Bovine serum albumin (BSA)
Trypsin/Lys-C (Promega V5073)
Dithiothreitol (DTT)
Iodoacetamide (IAM)
Trifluoroacetic acid (TFA)
Triethylammonium bicarbonate (TEAB)

4.2.1.2 Desalting

BiopureSPN Mini (Nest Group HUMS18V) desalting column.
Eppendorf Centrifuge 5417C.
Methanol washed 2 mL microcentrifuge tubes (ThermoFisher).

4.2.1.3 Tandem mass tag (TMT) labeling

Pierce Quantitative Colorimetric Peptide Assay

TMT Reagents (ThermoFisher)
DriSolv® Acetonitrile (Millipore AX01347)
50% (*v*/v) hydroxylamine (ThermoFisher 90,115)

4.2.1.4 Label-free

Peptide Retention Time Calibration Mixture (PRTC, Pierce 88,320, 0.5 pmol/μL)

4.2.1.5 LC-MS/MS

Isopropanol
Zorbax Chromatography Packing, SB-C18, 5 μm (Agilent)
UChrom C18 3-μm packing (NanoLCMS Solutions)
μM glass capillary (Agilent)
Model P-2000 Laser Puller (Sutter Instruments Co)
Capillary polishing station CPS-2 (EIS Source Solutions)
Pressure Injection Cell (NextAdance)
Easy-nLC 1200 (ThermoFisher)
Nanospray FlexIon Source (ThermoFisher)
Q Exactive Hybrid Quatrupole-Orbitrap Mass Spectrometer (ThermoFisher)

4.2.1.6 Data analysis

Proteome Discoverer (ThermoFisher)
Mascot search engine (Matrix Science)
Protein database in FASTA format (UniProt)

4.2.2 Protocols

4.2.2.1 Digestion

1. Reduction: Thaw the sample at room temperature. Dilute sample with 50 mM Tris, pH 8.0, to a concentration of 0.5 μg/μL in a volume of 100 μL, 25 μg total. BSA control: 50 μL (1 μg/μL), 50 μL 50 mM Tris, 50 μg total
2. Add 5 μL of 0.1 M DTT (to a final concentration of 5 mM), and mix
3. Incubate at 37 °C for 30 min. *Note: Avoid high temperatures (e.g., 50–60 °C) during the reducing step. High temperatures induce protein carbamylation in the presence of urea in the extract solution. Proteins are efficiently reduced at 37 °C*
4. Alkylation: Add 1.5 μL of 1 M IAM (to a final concentration of 15 mM), and mix

5. Incubate at room temperature for 30 min in the dark
6. Digestion: Add 400 μL of 50 mM Tris-HCl (pH 8). Dilution is needed to reduce extraction buffer concentration for optimal trypsin activity
7. Add enzyme to the desired enzyme to protein ratio. If using the trypsin/Lys-C mix, use 1:25 enzyme:protein ratio (2 μg trypsin/Lys-C:50 μg sample).
8. Incubate overnight at 37 °C in a Thermomixer C
9. Acidify sample by adding FA (5 μL) to a final concentration of 1% and remove particulate material by centrifuging at 14,000 × *g* for 10 min using a benchtop microcentrifuge. *Note: Acidification often results in cloudiness. Be sure to centrifuge the acidified sample*
10. Dry using the SpeedVac

4.2.2.2 Desalting

1. Reconstitute sample in 40 μL sample buffer (5% (*v*/v) acetonitrile/0.1% (v/v) TFA/water).
2. Desalt the sample following the manufacturer's protocol for the BioPure SPN Mini columns
3. SpeedVac to dryness and store at −20 °C until analysis on the LC-MS/MS (either label-free or TMT).

4.2.2.3 Tandem mass tag (TMT) labeling

1. Setup: Remove an aliquot from each sample and pool into a single tube. This will be used as a master control for TMT channel 11. Pool enough from each sample to allow for 10 μg of this control per 10-plex sample in the experiment. You will also need additional amount of the master control for testing digestion and labeling efficiency (i.e., if there are 100 samples, at least 110 μg of the master control will be needed).
2. Digest the master control and 25 μg of each sample (see Section 4.2.2.1).
3. Desalt the master control (see Section 4.2.2.2).
4. Reconstitute the master control to 1 μg/μL with 5% (*v*/v) acetonitrile/0.1% (v/v) FA/water
5. Transfer an aliquot of the master control (typically 1 μg) to a sample vial with insert and run on the LC-MS/MS
6. SpeedVac to dryness and store at −20 °C

   Note: If master control run on the LC-MS/MS looks good (proteins identified, high number of peptide spectrum matches), proceed with labeling.

7. Master control labeling: Reconstitute the master control in 100 mM TEAB to a concentration of 0.25 μg/μL
8. Remove 4 μL of master control to a new tube and add 16 μL distilled water (for protein concentration determination).
9. Determine the concentration using the Pierce Quantitative Colorimetric Peptide Assay kit
10. Label with TMT 131C according to the manufacturer's protocol
11. Transfer an aliquot of the master control (typically 1 μg) to a sample vial with insert and run on the LC-MS/MS. *Note: Analyze the master control labeled run to determine the labeling efficiency. If the labeling efficiency is good, proceed with sample labeling*
12. Sample labeling: Desalt the samples (see Section 4.2.2.2)
13. SpeedVac to dryness
14. Reconstitute the samples in 100 mM TEAB to a concentration of 0.25 μg/μL
15. Remove 4 μL of sample to a new tube and add 16 μL distilled water (for protein concentration determination).
16. Determine the concentration using the Pierce Quantitative Colorimetric Peptide Assay kit
17. Label each sample with TMT label according to the manufacturer's protocol
18. Pool 5 μg of each sample (labels 1–10) plus 5 μg of labeled master control (the volume is dependent upon the concentration of each sample that was determined in steps #9 and #16 above).
19. SpeedVac to dryness
20. Reconstitute TMT sample in 55 μL 5% (*v*/v) acetonitrile/0.1% (*v*/v) FA/water. This will give a final concentration of 1 μg/μL
21. Transfer to a sample vial with insert and run on the LC-MS/MS (typically a 1–5 μg injection).

4.2.2.4 Label-free

For label-free quantitative analysis, a known amount of Pierce Peptide Retention Time Calibration Mixture (PRTC, product #88320, 0.5 pmol/μL) will be spiked into the sample. The PRTC peptides can be used to normalize the data.

1. Digest (see Section 4.2.2.1) 25 μg of sample and desalt (see Section 4.2.2.2).
2. Add 23.75 μL 5% Buffer B (5% (*v*/v) acetonitrile/water/0.1% (v/v) FA) to the dried sample in the microcentrifuge tube. Final concentration of the sample will be 1 μg/μL

3. Add 1.25 μL PRTC (625 fmol) to the microcentrifuge tube. Final concentration of the PRTC will be 25 fmol/μL
4. Vortex and centrifuge
5. Transfer to a sample vial with insert for and run on the LC-MS/MS (typically a 1–5 μg injection).

4.2.2.5 LC-MS/MS

LC conditions (Easy nLC-1200)

- Column: Pulled glass emitter 75 μm × 20 cm capillary. The tip is packed with C18 5 μm material and the remainder of the column is packed with C18 3 μm material
- Buffer A: 0.1% (*v*/v) FA/water
- Buffer B: 80% (v/v) acetonitrile/0.1% (v/v) FA/water
- Wash solution 1 (WS1): isopropanol
- Wash solution 2 (WS2): 50% (v/v) acetonitrile/water
- Wash Solution 3 (WS3): 0.1% (v/v) FA/water
- Maximum pressure for all steps is 300 bar

Column equilibration: 10 μL**

Autosampler wash:

1 × 25 μL wash with WS3

3 × 25 μL washes with WS2

3 × 25 μL washes with WS

3 × 25 μL washes with WS2

3 × 25 μL washes with WS3

Gradient:

Both methods are run at 300 nL/min and incorporate a sawtooth gradient at the end of the run to facilitate column cleaning and reduce carryover.

Label-free gradient:

0–35% B in 80 min

35–70% B in 10 min

70–100% B in 3 min

3 sawtooth gradients from 2% to 100% B

TMT gradient:

0–35% B in 210 min

35–70% B in 10 min

70–100% B in 3 min

3 sawtooth gradients from 2% to 100% B

MS/MS settings (Q Exactive):

MS/MS is performed using a Q Exactive Hybrid Quatrupole-Orbitrap Mass Spectrometer coupled with a NanoSpray FlexIon Source using positive polarity, a spray voltage of 2.65 kV, capillary temperature of 250 °C and an S-Lens RF level of 50.

Label-free:

Full MS from 400 to 2000 m/z at 70,000 resolution with an AGC target of 5e5 and maximum inject time of 80 ms.

20 dd-MS2 at 17,500 resolution with an AGC target of 2e5 and maximum inject time of 80 ms. Isolation window of 2.0 m/z and NCE of 27.

Charge exclusion of +1, 6–8, >8, peptide match set to preferred and dynamic exclusion of 6.0 s.

TMT:

Full MS from 400 to 2000 m/z at 70,000 resolution with an AGC target of 1e6 and maximum inject time of 80 ms.

20 dd-MS2 at 35,000 resolution with an AGC target of 1e5 and maximum inject time of 50 ms. Fixed first mass of 110.0 m/z, isolation window of 1.2 m/z and NCE of 32.

Charge exclusion of +1, 6–8, >8, peptide match set to on and dynamic exclusion of 6.0 s.

After each sample run, a wash run is performed (3 sawtooth gradients), followed by 1–2 blank runs using a shortened gradient.

4.2.2.6 Data analysis

Data analysis is performed using Proteome Discoverer (PD; Thermo Fisher Scientific). Different workflows are used for the TMT and the label-free samples.

Label-free:

Mascot search against the database of interest with dynamic modifications of oxidation (Met) and deamidation (Asn, Gln) and static modification of carbamidomethyl (Cys). Sequest HT search against the PRTC peptides with static modification of heavy isotope labeled K and R. Normalization of data using the PRTC peptides. Precursor mass tolerance is set to 10 ppm and fragment mass tolerance is set to 0.02 Da.

After the runs for the sample set have been completed, create a study in PD. Add any study factors that may be used for comparing the samples and assign these factors to the samples. Add the .msf files for each of the samples to the study and create a new analysis comparing all the samples in the sample

set. Run a consensus workflow comparing all the samples. Run a consensus workflow with normalization and scaling on the PRTC peptides as well as imputation.

TMT:

Create a new quantitative method and add the reporter ion distributions for the TMT kit used. Mascot and Sequest HT search against the database of interest with dynamic modifications of oxidation (Met) and deamidation (Asn, Gln) and static modification of carbamidomethyl (Cys) and TMT (Lys, N-terminus). Precursor mass tolerance is set to 10 ppm and fragment mass tolerance is set to 0.02 Da.

After the runs for the sample set are complete, create a study in PD. Add any study factors that may be used for comparing the samples and assign these factors to the samples. Assign the samples in channel 131C as controls. Add the .msf files for each of the samples to the study and create a new analysis comparing all the samples in the sample set. Run a consensus workflow comparing all the samples. Run a consensus workflow with normalization and scaling on controls as well as imputation.

5. Conclusions

Here, we have presented two simple, inexpensive, and rapid methods to isolate low abundance seed proteins from soybean seeds. By employing these simple procedures, one can selectively remove most of the abundant seed storage proteins. Our methodology enables the nonabundant proteins in soybean seed to be more accessible for proteomic analysis. We have also provided detailed information on the resolution of low abundance proteins by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and the procedure for the identification of individual proteins by mass spectrometry. Procedure for TMT labeling is also detailed for the identification and quantification of proteins in different kinds of samples. The calcium chloride precipitation procedure could also be applied to other legumes to facilitate the enrichment of low abundance seed proteins. The methodology outlined here should help advance proteomic research of low abundance seed proteins of many legumes.

Acknowledgments

The authors would like to thank Nathan Oehrle for his excellent technical assistance. Funding for this research was provided by Agricultural Research Service, USDA (8042-21220-234-00D). Mention of a trademark, vendor, or proprietary product does not constitute a guarantee or warranty of the product by the USDA and does not imply its

approval to the exclusion of other products or vendors that may also be suitable. The US Department of Agriculture, Agricultural Research Service, Midwest Area, is an equal opportunity, affirmative action employer and all agency services are available without discrimination.

References

Agrawal, G. K., Hajduch, M., Graham, K., & Thelen, J. J. (2008). In-depth investigation of the soybean seed-filling proteome and comparison with a parallel study of rapeseed. *Plant Physiology*, *148*, 504–518.

Beilinson, V., Chen, Z., Shoemaker, R. C., Fischer, R. L., Goldberg, R. B., & Nielsen, N. C. (2002). Genomic organization of glycinin genes in soybean. *Theoretical and Applied Genetics*, *104*, 1132–1140.

Bourgeois, M., Jacquin, F., Savois, V., Sommerer, N., Labas, V., Henry, C., et al. (2009). Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. *Proteomics*, *9*, 254–271.

Cabanos, C., Matsuoka, Y., & Maruyama, N. (2021). Soybean proteins/peptides: A review on their importance, biosynthesis, vacuolar sorting, and accumulation in seeds. *Peptides*, *143*, 170598. https://doi.org/10.1016/j.peptides, 170598. Epub 2021 Jun 18. PMID: 34153351.

Cho, T. J., Davies, C. S., & Nielsen, N. C. (1989). Inheritance and organization of glycinin genes in soybean. *Plant Cell*, *1*, 329–337.

Derbyshire, E., Wright, D., & Boulter, D. (1976). Legumin and vicilin, storage proteins of legume seed. *Phytochemistry*, *15*, 3–24.

Hajduch, M., Ganapathy, A., Stein, J. W., & Thelen, J. J. (2005). A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. *Plant Physiology*, *137*, 1397–1419.

Islam, N., Bates, P. D., John, M. K., Krishnan, H. B., Zhang, Z., Luthria, D. L., et al. (2019). Quantitative proteomic analysis of low linolenic acid transgenic soybean reveals perturbations of fatty acid metabolic pathways. *Proteomics*, *19*, 1–11. https://doi.org/10.1002/pmic.201800379.

Islam, N., Krishnan, H. B., & Natarajan, S. S. (2020). Proteomic profiling of fast neutron-induced soybean mutant unveiled pathways associated with increased seed protein content. *Journal of Proteome Research*, *19*, 3936–3944. https://doi.org/10.1021/acs.jproteome.0c00160.

Islam, N., Krishnan, H. B., & Natarajan, S. S. (2021). Quantitative proteomic analyses reveal the dynamics of protein and amino acid accumulation during soybean seed development. *Proteomics*. https://doi.org/10.1002/pmic.202100143. Article e2100143.

Kim, Y. J., Wang, Y., Gupta, R., Kim, S. W., Min, C. W., Kim, Y. C., et al. (2015). Protamine sulfate precipitation method depletes abundant plant seed-storage proteins: A case study on legume plants. *Proteomics*, *15*, 1760–1764.

Kitamura, K., Takagi, T., & Shibasaki, K. (1976). Subunit structure of soybean 11S globulin. *Agricultural and Biological Chemistry*, *40*, 1837–1844.

Komatsu, S., Wang, X., Yin, X., Nanjo, Y., Ohyanagi, H., & Sakata, K. (2017). Integration of gel-based and gel-free proteomic data for functional analysis of proteins through soybean proteome database. *Journal of Proteomics*, *163*, 52–66.

Krishnan, H. B. (2000). Biochemistry and molecular biology of soybean seed storage proteins. *Journal of New Seeds*, *2*, 1–25.

Krishnan, H. B., Oehrle, N. W., & Natarajan, S. S. (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. https://doi.org/10.1002/pmic.200800875.

Li, C., & Zhang, Y. M. (2011). Molecular evolution of glycinin and beta-conglycinin gene families in soybean (Glycine max L. Merr.). *Heredity*, *106*, 633–641.
Low, T. Y., Syafruddin, S. E., Mohtar, M. A., Vellaichamy, A., Rahman, N. S., Pung, Y., et al. (2021). Recent progress in mass spectrometry-based strategies for elucidating protein–protein interactions. *Cellular and Molecular Life Sciences*, *78*, 5325–5339.
McWhite, C. D., Papoulas, O., Drew, K., Cox, R. M., June, V., Dong, X., et al. (2020). A pan-plant protein complex map reveals deep conservation and novel assemblies. *Cell*, *181*, 460–474.
Miernyk, J. A. (2014). Seed proteomics. *Methods in Molecular Biology*, *1072*, 361–377. https://doi.org/10.1007/978-1-62703-631-3_26 (PMID: 24136535).
Miernyk, J. A., & Hajduch, M. (2011). Seed proteomics. *Journal of Proteomics*, *74*, 389–400.
Miernyk, J. A., & Johnston, M. L. (2006). Chemical cross-linking immobilized-concanavalin A for use in proteomic analyses. *Preparative Biochemistry & Biotechnology*, *36*, 203–213.
Min, C. W., Hyeon, H., Gupta, R., Park, J., Cheon, Y. E., Lee, G. H., et al. (2020). Integrated proteomics and metabolomics analysis highlights correlative metabolite-protein networks in soybean seeds subjected to warm-water soaking. *Journal of Agricultural and Food Chemistry*, *68*, 8057–8067.
Min, C. W., Park, J., Bae, J. W., Agrawal, G. K., Rakwal, R., Kim, Y., et al. (2020). In-depth investigation of low-abundance proteins in matured and filling stages seeds of *Glycine max* employing a combination of protamine sulfate precipitation and TMT-based quantitative proteomic analysis. *Cell*, *9*(6), 1517. https://doi.org/10.3390/cells9061517.
Momma, T., Negoro, T., Hirano, H., Matsumoto, A., Udaka, K., & Fukazawa, C. (1985). Glycinin A5A4B3 mRNA: cDNA cloning and nucleotide sequencing of a splitting storage protein subunit of soybean. *European Journal of Biochemistry*, *149*, 491–496.
Mooney, B. P., Krishnan, H. B., & Thelen, J. J. (2004). High-throughput peptide mass fingerprinting of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. *Phytochemistry*, *65*, 1733–1744.
Natarajan, S. S., Krishnan, H. B., Lakshman, S., & Garrett, W. M. (2009). An efficient extraction method to enhance analysis of low abundant proteins from soybean seed. *Analytical Biochemistry*, *394*, 259–268.
Nielsen, N. C. (1996). Soybean seed composition. In D. P. S. Verma, & R. C. Shoemaker (Eds.), *Soybean: Genetics, molecular biology and biotechnology* (pp. 127–163). Tucson, AZ: CAB.
Nielsen, N. C., Dickinson, C. D., Cho, T., Thanh, V. H., Scallon, B. J., Fisher, R. L., et al. (1989). Characterization of glycinin gene family in soybean. *Plant Cell*, *1*, 313–328.
Ohyanagi, H., Sakata, K., & Komatsu, S. (2012). Soybean Proteome Database 2012: Update on the comprehensive data repository for soybean proteomics. *Frontiers in Plant Science*, *3*, 110.
Osborne, T. B. (1924). *The vegetable proteins*. London: Longmans, Green.
Shewry, P. R., Napier, J. A., & Tatham, A. S. (1995). Seed storage proteins: structures and biosynthesis. *Plant Cell*, *7*, 945–956.
Smolikova, G., Gorbach, D., Lukasheva, E., Mavropolo-Stolyarenko, G., Bilova, T., Soboleva, A., et al. (2020). Bringing new methods to the seed proteomics platform: Challenges and perspectives. *International Journal of Molecular Sciences*, *21*(23), 9162. https://doi.org/10.3390/ijms21239162. PMID: 33271881; PMCID: PMC7729594.
Staswick, P. E., Hermodson, M. A., & Nielsen, N. C. (1984). Identification of the cystines which link the acidic and basic components of the glycinin subunits. *Journal of Biological Chemistry*, *259*, 13431–13435.
Thanh, V., & Shibasaki, K. (1978). Major proteins of soybean seeds: subunit structure of β-conglycinin. *Journal of Agricultural and Food Chemistry*, *26*, 692–695.

van Wijk, K. J. (2001). Challenges and prospects of plant proteomics. *Plant Physiology*, *126*, 501–508.

Yu, T., Li, G., Dong, S., Liu, P., Zhang, J., & Zhao, B. (2016). Proteomic analysis of maize grain development using iTRAQ reveals temporal programs of diverse metabolic processes. *BMC Plant Biology*, *16*, 241.

Zecha, J., Gabriel, W., Spallek, R., Chang, Y., Mergner, J., Wilhelm, M., et al. (2022). Linking post-translational modifications and protein turnover by site-resolved protein turnover profiling. *Nature Communications*, *13*, 165. https://doi.org/10.1038/s41467-021-27639-0.

Zhang, M., Liu, S., Wang, Z., Yuan, Y., Zhang, Z., Liang, Q., et al. (2022). Progress in soybean functional genomics over the past decade. *Plant Biotechnology Journal*, *20*(2), 256–282. https://doi.org/10.1111/pbi.13682.