DOI 10.1002/pmic.200800875 *Proteomics* 2009, *9*, 3174–3188

RESEARCH ARTICLE

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*

Hari B. Krishnan[1,2], *Nathan W. Oehrle*[1] *and Savithiry S. Natarajan*[3]

[1] Plant Genetics Research Unit, Agricultural Research Service, United States Department of Agriculture, Columbia, MO, USA
[2] Division of Plant Sciences, University of Missouri, Columbia, MO, USA
[3] Soybean Genomics and Improvement Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD, USA

2-D analysis of plant proteomes containing thousands of proteins has limited dynamic resolution because only abundant proteins can be detected. Proteomic assessment of the non-abundant proteins within seeds is difficult when 60–80% is storage proteins. Resolution can be improved through sample fractionation using separation techniques based upon different physiological or biochemical principles. We have developed a fast and simple fractionation technique using 10 mM Ca^{2+} to precipitate soybean (*Glycine max*) seed storage globulins, glycinin and β-conglycinin. This method removes $87 \pm 4\%$ of the highly abundant seed proteins from the extract, allowing for 541 previously inconspicuous proteins present in soybean seed to be more detectable (volume increase of $\geq 50\%$) using fluorescent detection. Of those 541 enhanced spots, 197 increased more than 2.5-fold when visualized with Coomassie. The majority of those spots were isolated and identified using peptide mass fingerprinting. Fractionation also provided detection of 63 new phosphorylated protein spots and enhanced the visibility of 15 phosphorylated protein spots, using 2-D electrophoretic separation and an in-gel phosphoprotein stain. Application of this methodology toward other legumes, such as peanut, bean, pea, alfalfa and others, also containing high amounts of storage proteins, was examined, and is reported here.

Received: November 12, 2008
Revised: February 12, 2009
Accepted: March 1, 2009

Keywords:
β-Conglycinin / Glycinin / Legumes / Protein fractionation / Soybean

1 Introduction

The availability of genomic sequences has played an important role in establishing proteomics as a science. Progress in protein separation techniques, in conjunction with technological advances in MS, has allowed proteomics to develop quickly, where it is now playing an increasingly important role in genome annotation [1]. However, analysis of large proteomes consisting of sometimes thousands of proteins has limited dynamic resolution because only the most abundant proteins can be detected [2, 3].

Proteomic analysis of plant tissue is inherently difficult, requiring great attention to tissue isolation, handling and manipulation; and more importantly, protein isolation, preparation, storage and separation techniques. Isolation of all the different types of plant tissues for a proteomic analysis using standard, or even logically adapted protocols, can unintentionally reduce the amount of protein harvest or unknowingly introduce protein modifications [3–6], ulti-

Correspondence: Dr. Hari B. Krishnan, USDA-ARS, University of Missouri, 108 Curtis Hall, Columbia, MO 65211, USA
E-mail: Hari.Krishnan@usda.ars.gov
Fax: +1-573-884-7850

Abbreviations: C7BzO, 3-(4-heptyl)phenyl-3-hydroxypropyl)dimethylammoniopropanesulfonate; **2-HED,** 2-hydroxyethyl disulfide; **Pro-Q Diamond,** ProQ® Diamond fluorescent in-gel phosphoprotein stain (registered trademark of Molecular Probes); **SYPRO Ruby,** SYPRO® Ruby fluourescent in-gel protein stain (registered trademark of Molecular Probes)

mately making the final step, protein identification, difficult or impossible. Proteomic analysis of certain types of plant tissues, such as seed, leaf or tuber, is made even more difficult largely in part due to several or possibly many highly abundant proteins, thereby limiting the yield of those inconspicuous non-abundant proteins. For example, proteomic assessment of the non-abundant proteins within seed tissue, using general isolation and electrophoretic separation techniques, is difficult when the overwhelming majority, sometimes 60–80%, is made up of storage proteins [7].

However, dynamic resolution can be improved through fractionation of the proteome using separation techniques based upon different physiological or biochemical principles [3]. Development of a simple, fast and inexpensive method to remove the majority of storage proteins from a seed extract then would significantly enhance the study of the non-abundant proteins within seeds. Such a methodology would allow for those non-abundant proteins in seed to be assessed for (i) posttranslational modifications, such as glycosylation, phosphorylation, lipid-modification, processing and/or proteolysis [3], (ii) global relative protein quantity, mass and charge properties or immunoreactivity and (iii) complex formation and/or protein–protein interaction. Even more importantly, it would aid in the isolation of more mass-analyzable amounts of non-abundant seed proteins; allowing for the discovery of new or novel proteins within the seed proteome, or simply verification of identity, *via* one of the many different types of massspectrometers available today.

Development of such fractionation techniques for commercial seed crops, such as soybean, would be highly useful, allowing more clues about their nutritive value, yield potential and environmental stress responses to be ascertained. To meet the challenges of future seed crop demands, proteins in control of plant architecture, metabolism and stress response/resistance will need to be identified to facilitate higher crop productivity through biotechnological improvement [7]. For example, seed cultivation for soybean was 230 million metric tons for 2006, providing 68% of the consumable protein meal worldwide (US Department of Agriculture, 2007). Hence, soy seed research has focused great attention toward seed protein quality improvement. Demand for quality soy seed protein and its products has increased steadily for decades and is not expected to decline. Unfortunately, analysis of any improvements to soy seed proteins made through many types of biotechnological manipulation, or even breeding, is made difficult because cultivated soy total seed protein consists mainly, sometimes 80% or higher, of storage proteins [8].

Almost 80 storage proteins have been identified in soy seed [9] and each fall into one of the four basic categories: albumins (water-soluble), globulins (salt-soluble), prolamins (alcohol-soluble), and glutelins (weak acid/weak base-soluble) [10]. Cultivated soybean (*Glycine max*) seed storage proteins consist primarily of two major storage protein complexes, glycinin and β-conglycinin [11–13], which fall into the globulin category (soluble in dilute aqueous-salt solutions). The other seed storage proteins only account for a minor portion of the total seed protein content since globulins can be 60–80% of the total seed protein. Interestingly, of these two major seed storage protein complexes, substantial portions of their subunits are allergenic [14–17].

Glycinin, accounting for roughly 40–60% of the total seed protein, is a hexameric protein, ranging from 320 to 375 kDa. It is composed of the G1, G2, G3, G4 and G5 subunits (approximately 56, 54, 54, 64 and 58 kDa, respectively), all of which consist of one acidic and one basic chain (approximately 37–44 and 17–22 kDa, respectively) [11, 18, 19]. β-Conglycinin, accounting for roughly 30–40% of the total seed protein, is a trimeric glycoprotein, ranging from 126 to 170 kDa. It is 5% carbohydrate by weight and composed of three subunits, α, α′ and β (approximately 76, 72 and 53 kDa, respectively) [8, 20, 21].

Isolation, or removal, of some of the globulins from the salt-soluble fraction can be accomplished through column chromatography [22, 23], manipulation of solvent properties [24–28] or a combination of techniques [13, 29, 30]. Inherent difficulties are encountered with these techniques however: cross-reactivity, broad range affinity, low reproducibility and expense, to name a few. In addition, some of these methods can remove other proteins not intended upon as targets for removal. From a proteomic perspective, loss of too much or all of a certain species of protein can make quantitative and qualitative analysis difficult or impossible. Even more importantly, these procedures can take extensive amounts of time to perform, running the risk of proteins of interest, or possibly a specific target protein, being modified or degraded if the fractionation procedure cannot be done under denaturing conditions or at low temperature to prevent proteolysis.

For some time, it has been known that Ca^{2+} binds to soybean glycinin [31] and β-conglycinin [32], and depending on solution pH, ionic strength, temperature and calcium concentration can cause both associative inter- and intra-aggregation among the globulins [10, 33, 34]. Building upon this knowledge of soy globulins and Ca^{2+} association, Deak *et al.* [35] sought a simple method using small amounts of calcium salts as fractionation agents. With this knowledge, our approach was to simplify even further the removal of the vast majority of globulins from the salt-soluble fraction, in an effort to have a fast and effective way of reducing or possibly excluding them from our protein preparations for 2-D electrophoretic separation and analysis. The results of this simple Ca^{2+}-fractionation step and electrophoretic analysis with soybean seed, carried out with 1-D and the more complex 2-D separation, are reported here.

In addition, success in utilizing such an approach with cultivated soybean seed brought us to hypothesize that the same or a similar technique might be universally applicable to other economically important crop seed legumes, since legume globulin homology has been shown to be highly conserved [10, 36]. Such a methodology would be very beneficial in the area of legume research, since many are struggling with the same issues with seed storage proteins. Those results,

using the same simplified approach, but with peanut, beans, peas, alfalfa, vetch and others, are also reported here.

2 Material and methods

2.1 Reagents

Calcium chloride dihydrate, magnesium chloride hexahydrate, manganese chloride tetrahydrate, urea (electrophoresis grade), thiourea (electrophoresis grade), methanol (HPLC grade), ammonium acetate, β-mercaptoethanol, glycerol, mineral oil, agarose (low EEO) and buffer reagents were obtained from Fisher Scientific (Pittsburgh, PA, USA). Phenol (Tris-equilibrated), CHAPS, 3-(4-heptyl)phenyl-3-hydroxypropyl)dimethylammoniopropanesulfonate (C7BzO), 2-hydroxyethyl disulfide (2-HED), DTT, iodoacetamide and protease inhibitor cocktail (P-9599) were obtained from Sigma-Aldrich (St. Louis, MO, USA). Acrylamide, bis-acrylamide, ammonium persulfate, TEMED, and CBB G-250 were obtained from BioRad (BioRad Laboratories, Hercules, CA, USA). Pro-Q Diamond phosphoprotein stain (ProQ® Diamond fluorescent in-gel phosphoprotein stain (registered trademark of Molecular Probes)) and SYPRO Ruby (SYPRO® Ruby fluourescent in-gel protein stain (registered trademark of Molecular Probes)) fluorescent protein stain was obtained from Molecular Probes (Invitrogen, Carlsbad, CA, USA). IPG strips, IPG buffer 4–7 and 3–10, Cy3 and Cy5 fluorescent dyes were obtained from GE Healthcare (Piscataway, NJ, USA).

2.2 Plant material

Fractionation experiments included salt-soluble proteins extracted from the following legumes: soybean [*G. max* (L.) Merr. cv Williams 82], peanut [*Arachis hyogaea* (L.)], pigeon pea [*Cajanus cajan* (L.) Huth.], chick-pea [*Cicer arietinum* (L.)], Alfalfa [*Medicago sativa* (L.)], Smooth Vetch [*Vicia dasycarpa* (Ten.)], Adzuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi], black gram [*Vigna mungo* (L.) Hepper.], mung bean [*Vigna radiate* (L.)], cowpea [*Vigna unguiculata* (L.) Walp.], common bean [*Phaseolus vulgaris* (L.) cv. Kentucky Wonder Pole], Lima bush bean [*Phaseolus lunatus* (L.)], garden Lupin [*Lupinus polyphyllus* (Lindl.)], Bird's-foot Trefoil [*Lotus corniculatus* (L.)], fenugreek [*Trigonella foenum-graecum* (L.)], and American potato [*Apios americana* (Medik.)].

2.3 Protein extraction and fractionation

Mature, dried soybean seeds were ground into a fine powder using a chilled mortar and pestle and 20 mg weighed and placed into a tube for extraction using 1 mL of 20 mM Tris-HCl, pH 6.8, containing protease inhibitor cocktail. Protein extraction was carried out for 15 min, with vigorous shaking, at room temperature. Solution was clarified with centrifugation at $16\,100 \times g$ for 10 min. Supernatant was removed and placed into a clean tube for fractionation. All other legume species seed samples were mature, dried seed, also ground into a fine powder using a chilled mortar and pestle. However, 40 mg were weighed and placed into a tube for extraction using 1 mL of 20 mM Tris-HCl, pH 6.8. Similar protein extraction methods were carried out: 15 min, with vigorous shaking, at room temperature followed by clarification. For fractionation, a 100 mM $CaCl_2 \cdot 2H_20$ stock solution was prepared and used to bring a sample of each extraction to the final concentration of calcium chloride desired. Similar stock solutions of $MgCl_2 \cdot 6H_20$ and $MnCl_2 \cdot 4H_20$ were made for alternate fractionation trials. After addition of the fractionation agent, and an additional 10 min of shaking, each solution was clarified with centrifugation at $16\,100 \times g$ for 10 min. The supernatant was removed, placed into a clean tube and placed immediately on ice.

2.4 1-D electrophoresis

For 1-D analysis the clarified supernatant was combined 1:1 with 2 × SDS-PAGE sample buffer. 1-D separation followed the method of Laemmli [37] using 13.5%T gels run using Mini250 (GE Healthcare). Separation was achieved with a constant 20 mA *per* gel and run time of 1.5 h. Gels were removed from the cassette and placed immediately in Coomassie staining solution (20% ethanol, 8% ammonium sulfate, 1.6% phosphoric acid, 0.4 g/L CBB G-250). Typically, 20–40 μg of protein from each sample was loaded *per* well.

2.5 2-D electrophoresis

After the initial extraction and fractionation, the clarified supernatant was placed on ice and quickly brought to 0.9 M sucrose, adjusted to 100 mM Tris-HCl, pH 8.8 with 1.5 M Tris-HCl, pH 8.8 and β-mercaptoethanol added to 0.4%. An equal volume of Tris-equilibrated phenol was then added with substantial vortexing. Each sample was then mixed vigorously for 30 min at 22°C followed immediately by centrifugation at $4000 \times g$ for 20 min at 15°C in a swing-bucket rotor. The upper phenolic phase was removed and added to five volumes of freshly prepared 100% methanol with 0.1 M ammonium acetate (chilled to −80°C). Protein precipitation progressed for 2 h at −80°C and was followed by centrifugation at $6000 \times g$ for 10 min at 4°C. The supernatant was discarded and protein pellet was resuspended vigorously in a freshly prepared solution of 100% methanol with 0.1 M ammonium acetate and 10 mM DTT (chilled to −20°C). Washing of the insoluble proteins was repeated three times with the same solution with incubation at −20°C for 20 min followed by centrifugation at $12\,000 \times g$

for 10 min at 4°C in between each vigorous wash step. Washing of the insoluble proteins was repeated twice more with a freshly prepared solution of 80% acetone containing 10 mM DTT with incubation at −20°C for 20 min followed by centrifugation at 12 000 × *g* for 10 min at 4°C in between each vigorous wash step. For the final wash of the proteins, 80% ethanol with 10 mM DTT (freshly prepared and chilled to −20°C) was used. Following a hard centrifugation as with the washes, the protein pellet was allowed to air dry to near dryness. Proteins were then solubilized in a small volume of 7 M urea, 2 M thiourea, 1% CHAPS, 2% C7BzO and 100 mM DTT with vortexing. Samples were then stored on ice, never frozen, until protein concentration obtained and samples diluted for IEF.

For standard IEF, protein estimation was performed following the method of Bradford [38] and 200 μg of protein sample loaded *per* strip using in-gel rehydration. Linear gradient, 13 cm IPG strips (GE Healthcare) were brought to a rehydration volume of 250 μL with 7 M urea, 2 M thiourea, 1% CHAPS and 2% C7BzO with a final concentration of the following: 5% glycerol, 2.2% 2-HED and 0.25% 4–7 IPG buffer or 0.5% 3–10 IPG buffer. The final concentration of DTT in each sample load, optimized previously to be used in conjunction with 2-HED [39], was adjusted to precisely to 60 mM. Rehydration solutions were vortexed with moderate force and incubated on ice for 30 min. Strips were then passively rehydrated with the entire rehydration solution containing protein sample at 22°C for 15 h prior to focusing.

IEF method was as follows: 50 V active rehydration, 1 h; 250 V, 250 V h, fast ramp; 1000 V, 500 V h, fast ramp; 8000 V, 2 h, linear ramp; 8000 V, 40–60 000 V h, fast ramp and was performed with a Protean II IEF (BioRad). Separation was finalized when current was at or below 20 μA *per* strip and steady for 1 h. Prior to the second dimension, IPG strips were equilibrated with 5% SDS in a urea-based solution (50 mM Tris-HCl, pH 8.8, 6 M urea, 30% glycerol and 0.1% bromophenol blue) containing 2% DTT for 20 min and again but with 2.5% iodoacetamide for 20 min. IPG strips were carefully placed onto a Hoeffer SE600 (GE Healthcare) 16%T vertical second dimension and secured into place with warm 1% agarose dissolved in SDS-PAGE running buffer (0.2% SDS). Gels were run at an initial 10 mA/gel for 1 h followed by 30 mA/gel for the remainder of the run (elimination of dye front; approximately 3.5 h). 2-D gels for non-fluorescent staining were immediately removed and fixed in 5:4:1 (methanol:water:acetic acid) for 1 h, followed by staining in Coomassie G-250 for 24 h. 2-D gels for in-gel phosphoprotein analysis were immediately removed and fixed in 5:4:1 (methanol:water:acetic acid) for overnight (with two changes of fixative), followed by staining in Pro-Q for 3 h and post-staining with SYPRO Ruby for overnight.

2.6 DIGE

DIGE was performed identical to previous 2-D methodology, however (i) the initial protein isolate was solubilized in 7 M urea, 2 M thiourea, 1% CHAPS and 2% C7BzO and 30 mM Tris-HCl, pH 8.5 (without reducing agent) prior to dye labeling and (ii) the exact protein concentration was determined using the EZQ system (Molecular Probes). For DIGE, a linear gradient 13 cm, 4–7 IPG was utilized and 50 μg of each protein sample was labeled with 400 pmol of both Cy3 and Cy5 (separately) following the manufacturer's protocol (GE Healthcare). Exactly 100 μg was loaded using in-gel rehydration on each strip after combining a Cy3-labeled sample and a Cy5-labeled sample of both (dye swap) and bringing the final rehydration volume to 250 μL with 7 M urea, 2 M thiourea, 1% CHAPS and 2% C7BzO with a final concentration of the following: 5% glycerol, 2.2% 2-HED and 0.25% 4–7 IPG buffer. Again, the final concentration of DTT in each strip, optimized previously to be used in conjunction with 2-HED [39], was adjusted precisely to 60 mM. Rehydration solutions were vortexed with moderate force and incubated on ice for 30 min prior to loading. Strips were then passively rehydrated with entire rehydration solution containing protein sample at 22°C for 15 h protected from light prior to focusing. After 2-D separation, CyDye-labeled DIGE gels were immediately removed from the cassette and soaked in 20% methanol for 1 h, protected from direct light, prior to imaging.

2.7 Phosphoprotein and SYPRO Ruby analysis

In-gel phosphorylated protein spot analysis was performed identical to previous 2-D methodology, however the exact protein concentration was determined using the EZQ system. For phosphoprotein analysis, 200 μg of each protein sample was run *per* gel and after electrophoresis, gels were processed with a modified Pro-Q Diamond methodology [40]. After imaging, gels were briefly rinsed with ultrapure water, and stained for total protein using SYPRO Ruby and processed according to the manufacturer's protocol prior to imaging.

2.8 Image acquisition and analysis

DIGE gels were scanned using an *Ettan*DIGE Imager v1.0, processed using ImageQuant TL and analyzed for proteome differences using DeCyder Differential Image Analysis v5.01 (GE Healthcare) and Delta2D v3.6 (Decodon, Greifswald, Germany). Gels stained with Pro-Q Diamond were scanned using a Fuji FLA5000 v3.0 at 532 nm excitation with a 575 long-pass green filter, and optimized for image quality using Fuji Multi Gauge v2.3. In-gel phosphorylated protein-stained gels post-stained with SYPRO Ruby were scanned using a Fuji FLA5000 at 473 nm excitation with a 510 nm long-pass blue filter, and optimized for image quality using Fuji Multi Gauge. Images from ProQ Diamond-stained gels and images from SYPRO Ruby-stained gels were analyzed for proteome differences using

Table 1. Proteins identified from soybean seed protein extract after using Ca^{2+} fractionation method, 2-DE separation and peptide mass fingerprinting (MALDI-TOF MS)

SID	Protein identification	MOWSE	Peptides matched	% Sequence coverage	NCBI accession	Fold reduction (control/Ca^{2+})	$\pm$ ($n = 4$)
1	β-Conglycinin (α-subunit)	267	31	53	gil9967357	18.91	4.28
2	β-Conglycinin (α′-subunit)	282	31	48	gil9967361	10.06	1.55
3	Sucrose-binding protein precursor	63	13	28	gil548900	2.05	0.06
4	Sucrose-binding protein precursor	61	12	28	gil548900	2.14	0.05
5	Sucrose-binding protein precursor	109	14	28	gil548900	3.35	0.41
6	Sucrose-binding protein precursor	113	16	31	gil548900	4.04	0.52
7	Sucrose-binding protein precursor	83	16	30	gil548900	5.32	0.26
8	β-Conglycinin (β-subunit)	188	19	43	gil63852207	6.94	1.34
9	β-Conglycinin (β-subunit)	194	25	49	gil21465628	8.64	1.60
10	β-Conglycinin (β-subunit)	255	26	52	gil21465628	12.17	2.05
11	β-Conglycinin (β-subunit)	260	29	52	gil21465628	10.91	1.92
12	β-Conglycinin (β-subunit)	*63400*	28	52	gil63852207	11.58	0.72
13	β-Conglycinin (β-subunit)	247	28	49	gil21465628	9.28	1.06
14	β-Conglycinin (α-subunit)	143	19	33	gil9967357	3.66	0.37
15	β-Conglycinin (α-subunit)	142	20	36	gil9967357	6.14	0.57
16	β-Conglycinin (α-subunit)	215	21	40	gil9967357	5.79	1.15
17	Glycinin (A3B4)	76	11	22	gil33357661	9.68	2.86
18	Glycinin (A3B4)	86	10	22	gil33357661	13.17	3.63
19	Glycinin (A3B4)	70	9	18	gil33357661	12.43	2.93
20	Glycinin	126	10	39	gil6015515	13.71	5.67
21	Glycinin	122	10	39	gil6015515	102.57	23.34
22	Glycinin (A2B1a)	96	10	21	gil121277	9.99	2.02
23	Glycinin (A2B1a)	102	10	21	gil121277	16.79	3.70
24	Glycinin (A2B1a)	72	10	23	gil121277	11.75	3.25
25	Glycinin (A1aBx) precursor	63	8	18	gil121276	12.00	3.74
26	Proglycinin (A1ab1b)	76	9	19	gil15988117	11.17	2.85
27	Proglycinin (A1ab1b)	79	8	19	gil15988117	9.87	2.00
28	Glycinin (A1aBx) precursor	63	9	20	gil121276	14.45	5.32
29	Proglycinin (A1ab1b)	96	10	20	gil15988117	9.72	1.69
30	Glycinin (A1aBx) precursor	204	7	16	gil121276	7.60	0.90
31	Glycinin	96	9	39	gil6015515	6.82	0.77
32	Glycinin	*77868*	5	22	gil6015515	4.80	0.75
33	Glycinin (A2B1a) precursor	60	4	26	gil169967	22.72	4.67
34	Glycinin (A1aBx) precursor	*102300*	7	17	gil121276	5.82	1.40
35	Glycinin (A2B1a) precursor	60	5	27	gil169967	26.28	7.23
36	Glycinin (A1aBx) precursor	61	7	27	gil121276	29.46	7.53
37	Glycinin (A1aBx) precursor	*53758*	7	13	gil121276	15.32	3.94
38	Glycinin (A5A4B3) precursor	*737*	7	8	gil121279	22.32	1.33

These 38 proteins consist primarily of the many subunits of the seed storage globulins, glycinin and β-conglycinin. After 10 mM Ca^{2+} fractionation, they are reduced in %volume by $87 \pm 4\%$, to comprise less than 6% of the total protein. Average (weighted) fold reduction data were generated with Delta2D image analysis software, using images from three separate fractionations, four separate 2-DE and visualization techniques (DIGEx2/SYPRO Ruby/Coomassie). Average was calculated using weight ratios calculated from relative precision of imaging method. Error is reported as the weighted standard deviation of the average fold reduction ($n = 4$). Mascot ''probability-based MOWSE scores'' ≥ 57 represent those above 95% confidence limit ($p \leq 0.05$). MOWSE scores in italics represent those searches performed *via* Protein Prospector. Searches were confined to *Glycine max* databases using a peptide mass tolerance of 20 ppm. Spot # identifier (SID) corresponds to those shown in Fig. 4A.

Delta2D. Coomassie-stained gels were destained with multiple changes of ultrapure H_2O to remove background and scanned using a HP Scanjet 5470c controlled through Adobe Photoshop. DeCyder analysis software was set to default parameters for normalization using maximum volume and a differential detection threshold of twofold using spot volume ratio. DeCyder was primarily used for differential image acquisition (two channel overlay) while Delta2D provided %volume data used throughout this report.

Delta2D parameters were set to maximize spot detection (using *global* image warping and *exact* spot matching), hence very little background subtraction was used throughout. Percent spot volume ratio (master gel [control, no calcium fractionation]/experimental gel [calcium fractionated]) differences were noted in the spot quantitation table and cutoffs were determined according to fold changes, keeping nearly all protein spots for calculation/comparison purposes. Here, the relative quantity of the spot, excluding background gave a more accurate determination of the %change since the total quantity of all spots on the gel is 100%. Average fold reduction data included in Table 1 was calculated from four independent image comparisons (control *versus* fractionated) executed in Delta2D, generated from three independent fractionations and the three different visualization techniques used (DIGE1, DIGE2, SYPRO Ruby and Coomassie). Spot% volume ratio [control]/[calcium fractionated] differences were calculated within each comparison ($n = 4$) and the reciprocal (fold reduction) was calculated using weight ratios calculated from relative precision of each imaging method (*i.e.* less ratio variance = higher weighted ratio). Error is reported as the weighted standard deviation of the average fold reduction.

2.9 Protein identification

A small gel piece of each protein for identification was excised with a 1.5 mm Spot Picker (The Gel, San Francisco, CA, USA) from a Coomassie G-250-stained gel, washed briefly in distilled water and then destained completely in a 50% v/v solution of ACN containing 25 mM ammonium bicarbonate. After a 100% ACN wash, the protein contained in the acrylamide gel was subjected to digestion using 20 μL (10 μg/mL) of modified porcine trypsin (Promega, Madison, WI, USA) in 25 mM ammonium bicarbonate. Peptides resulting from the tryptic digestion were analyzed using a Voyager DE-STR (Applied Biosystems, Framingham, MA, USA) MALDI-TOF mass spectrometer. The peptides were co-crystallized with CHCA matrix on a MALDI-TOF MS plate, briefly dried and ionized using a 337 nm nitrogen laser operating at 20 Hz. Trypsin autolysis peaks of charge/mass ratios 842.51, 1045.56 and 2211.10 served as internal calibrants. Peptide mass searches were performed *via* MASCOT (Matrix Sciences, http://www.matrixscience.com) and Protein Prospector (University of California, San Francisco, http://prospector.ucsf.edu) using primarily the NCBI (National Center for Biotechnological Information) non-redundant *G. max* database and UniProtKB *G. max* database. All searches were performed with a fragment mass tolerance of 20 ppm, allowance of only one missed cleavage, and carbamidomethyl fixed modification. Peptides from unmatched proteins were searched using identical search criteria, but within ''all-plant'' databases.

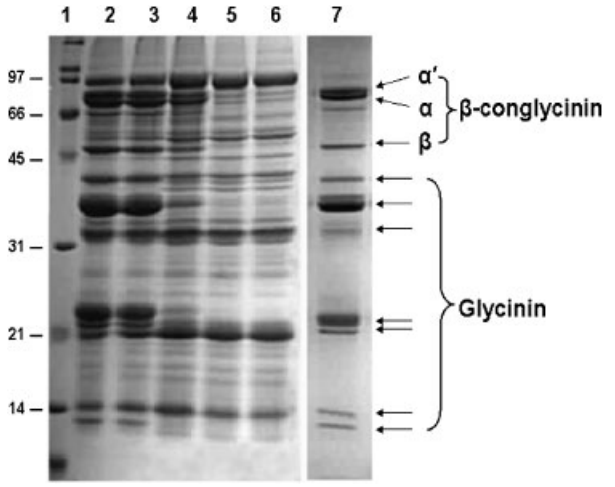

Figure 1. Fractionation of soybean (cv. Williams 82) seed proteins with increasing Ca^{2+}. Salt-soluble proteins were extracted from soybean seed using a low ionic strength buffer and the effects of increasing calcium chloride on protein fractionation were examined. Lane 1, protein molecular weight markers in kDa; Lane 2, no calcium added; Lane 3, 1 mM $CaCl_2$; Lane 4, 2 mM $CaCl_2$; Lane 5, 5 mM $CaCl_2$; Lane 6, 10 mM $CaCl_2$; Lane 7, resulting precipitant from calcium chloride fractionation, demonstrating the removal of almost exclusively β-conglycinin and glycinin.

3 Results and discussion

3.1 Fractionation and 1-D analysis

Fractionation of the salt-soluble proteins was performed using calcium chloride as outlined in Section 2. 1-DE results revealed a significant reduction in the amount of recovered protein in the salt-soluble fraction with increasing amount of added Ca^{2+} (Fig. 1). Fractionation differences between the 2 and 5 mM Ca^{2+} revealed significant reduction in several highly abundant proteins. At first glance, fractionation using greater than 5 mM Ca^{2+} appeared to be unnecessary, but additional analysis revealed that 10 mM Ca^{2+} was more effective at near complete removal of approximately ten protein bands that comprised a significant portion of the total protein extract. A repeat of the 10 mM calcium chloride fractionation was performed, but in addition to the soluble fraction, an aliquot from those precipitated proteins was analyzed using 1-DE. Precipitated proteins were dissolved in SDS-PAGE sample buffer and according to the protein concentration, adjusted volumetrically to achieve an equal protein load for 1-DE analysis. The resulting separation revealed a near complete removal of those highly abundant proteins within soybean seed extract (confirmed later to consist primarily of soybean seed storage globulins), with only minor losses of proteins not targeted for removal (Fig. 1). Also revealed, with an equal protein load of the resulting soluble fraction, was the now more visible non-abundant proteins within soybean seed.

The divalent cations Mg^{2+} and Mn^{2+} were used in place of Ca^{2+} at an equal concentration in an attempt to possibly eliminate the removal of non-target proteins. The results using 1-DE revealed a reduction in the amount of globulins remaining in solution using all three of the divalent cations (data not shown), however Mn^{2+} precipitated higher amounts of non-target proteins of interest and Mg^{2+} did not force the precipitation of either α or α′ subunits of β-conglycinin, nor completely eliminate the β subunit of β-conglycinin. Both Mg^{2+} and Mn^{2+} did fractionate the acidic and basic subunits of glycinin, hence either of these could possibly be used in an alternate protocol for fractionation of glycinin alone. The results also demonstrated some slight differences in the species of non-fractionated proteins retained in the soluble fraction depending on the divalent cation used.

3.2 2-DE separation and analysis

Our goal for using Ca^{2+} fractionation was to have a sample enriched for the non-abundant proteins in soybean seed, through elimination of the highly abundant seed storage proteins. Ultimately, we wanted to discover the inconspicuous proteins in soybean seed using 2-DE separation and have another tool for verifying proteome changes resulting from genetic perturbation. To test if this methodology was truly an effective approach, we first employed the technique of DIGE. This technique eliminates the variability that can be seen when trying to compare two samples, run in parallel on two separate gels. Two freshly prepared seed extracts, as prepared previously, one fractionated with 10 mM Ca^{2+} the other not, were labeled with both Cy3 and Cy5 (individually), then combined, and 2-DE separated. The resulting DIGE image overlays (Fig. 2A, red spots) clearly demonstrate a large reduction in the amount of those highly abundant proteins normally seen on 2-DE separated soybean seed extract, with an obvious increase in those lower abundance seed proteins (Fig. 2A, green spots). Duplication of the electrophoretic separation using the same samples, where each sample was labeled with both Cy3 and Cy5 but combined in reverse manner for two separate parallel gel runs (Fig. 2B), revealed the same results since previously what was red is now green, and what was green is now red. Most importantly, either DIGE image reveals the enhancement of non-abundant proteins in the fractionated sample; since little or no software generated yellow spots (overlapping *and* equally abundant) appear.

DIGE image comparison using Delta2D image analysis software revealed that of the 1484 spots detected, 38 of those spots decreased more than 2.5-fold in %volume (range 2.700–248, mean 53.411) and 293 spots increased more than threefold in %volume in the calcium fractionated sample (range 3.014–90.688, mean 14.165). These 38 proteins

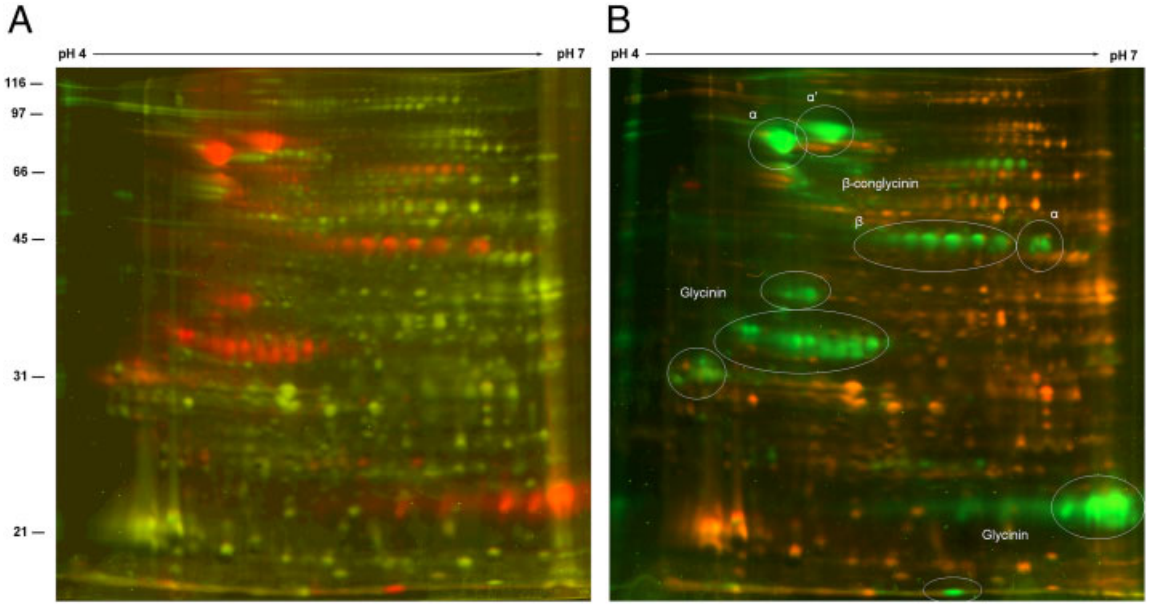

Figure 2. DIGE analysis of 10 mM Ca^{2+} fractionated soybean (cv. Williams 82) seed proteins using CyDyes. (A) Overlay image of the non-fractionated sample (labeled with Cy3; red) and fractionated sample (labeled with Cy5; green). (B) Overlay image of the non-fractionated sample (labeled with Cy5; green) and fractionated sample (labeled with Cy3; red). Clearly the non-abundant seed proteins, normally inconspicuous due to the overabundance of seed storage proteins, are enhanced. The resulting % volume data generated from these images using Delta2D image analysis software revealed that of the 1484 spots detected, 38 spots decreased more than 2.5-fold in %volume and 293 spots increased more than threefold in % volume in the calcium fractionated sample. Labels in Panel (B) highlight the proteins removed *via* fractionation (bright green). Corresponding proteins are unlabeled in (A) (red) for image clarity. IEF separation was from pH 4 to 7 and protein molecular weight markers are in kDa.

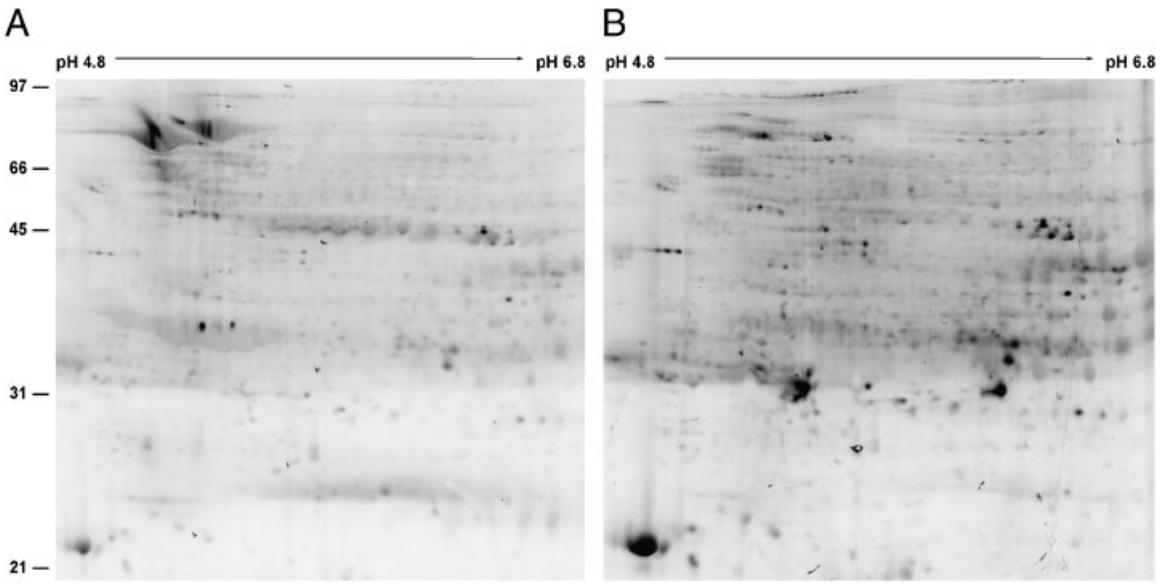

Figure 3. 2-DE separation and analysis using ProQ Diamond in-gel phosphorylated protein stain. Non-fractionated soybean seed proteins (A) and 10 mM Ca^{2+} fractionated soybean seed proteins (B) were analyzed for the presence of phosphorylated protein spots using ProQ Diamond in-gel stain. The appearance and/or enhancement of signal from those non-abundant proteins usually masked by those storage proteins (which limit IEF strip load and interfere with 2-D separation) is clearly evident. Many previously undetected phosphorylated protein spots (63 spots) are visible now, while other phosphorylated protein spots (15 spots) have enhanced signal making them easier to discern from background. Data were generated using Delta2D image analysis software. IEF was from pH 4 to 7 and protein molecular weight markers are in kDa.

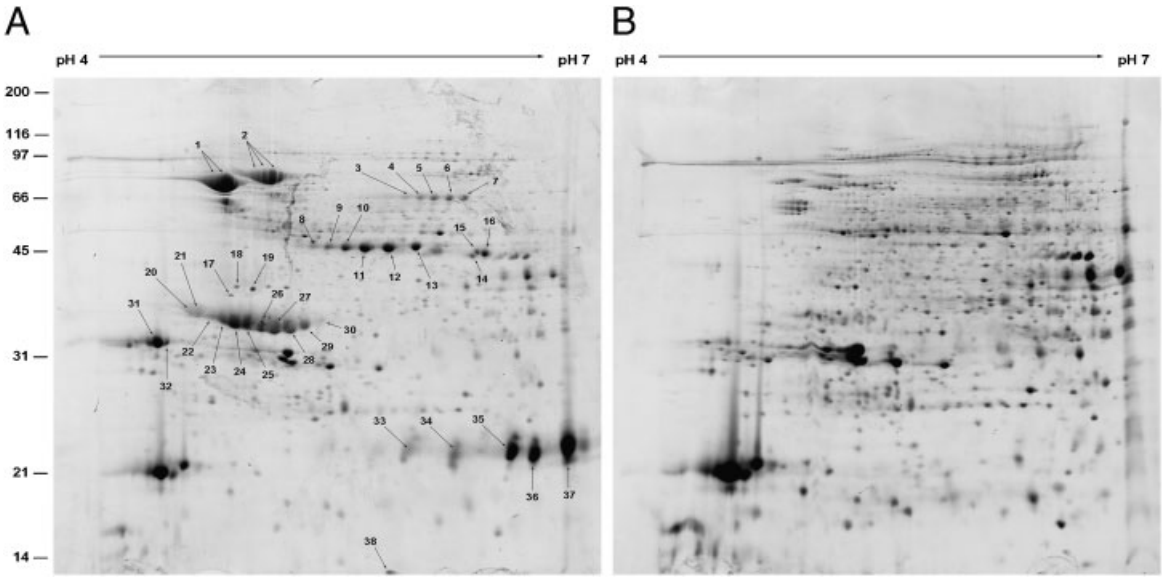

Figure 4. 2-DE analysis of non-fractionated soybean seed proteins (A) and 10 mM Ca^{2+} fractionated soybean seed proteins (B) detected with SYPRO Ruby in-gel fluorescent protein stain. Those gels previously stained with ProQ Diamond (Fig. 3) were post-stained with SYPRO Ruby after phosphoprotein analysis. Similar to the DIGE analysis (Fig. 2), a massive enhancement in those non-abundant proteins can be seen using Ca^{2+} fractionation. Here, the non-fractionated and fractionated samples (equal protein load) were run on two separate gels to demonstrate the effectiveness of the method for spot-picking and possible peptide mass fingerprinting. Of the 1184 total spots detected, 293 new spots were additionally detectable after calcium fractionation (891 total spots detected in control image). Protein spot numbers in (A) correspond to those listed in Table 1. IEF separation was from pH 4 to 7 and protein molecular weight markers are in kDa.

removed *via* Ca^{2+} fractionation had previously comprised more than 50% of the total separated protein, but now comprised less than 6%. Comparatively, that was a 90% reduction of those 38 proteins. Peptide mass fingerprinting using MALDI-TOF MS subsequently done on those 38 protein spots taken from a similar electrophoretic separation was used to identify these proteins (Table 1). MS data collected revealed that the overwhelming majority of those depleted proteins were composed of the soybean seed storage globulins, β-conglycinin and glycinin. More specifically, 98% (*per* volume) of those proteins removed *via* calcium fractionation were comprised of α, α′ and β subunits of β-conglycinin and the acidic and basic subunits of glycinin. The remaining 2% (*per* volume) of proteins removed consisted of the somewhat abundant sucrose-binding protein.

Additional identically prepared samples were 2-DE separated within the same range (pH 4–7) and analyzed for the presence of phosphoproteins using ProQ Diamond in-gel phosphorylated protein stain (Fig. 3). This was an effort to see if either the appearance or enhancement of signal from those non-abundant phosphoproteins, possibly masked by those storage proteins which limit IEF strip load and interfere with 2-DE separation, could be detected. Previously undetectable phosphorylated protein spots (68 spots) are visible after fractionation, while those previously difficult to detect phosphorylated protein spots (15 spots) have enhanced signal making them easier to discern from background. Calcium fractionation did somewhat deplete 11 protein spots detectable using ProQ Diamond, however these proteins comprise less than 1% of the total number of proteins detected. While fractionation would interfere with actual global quantization of protein phosphorylation, it clearly provides a method for detecting novel proteins using specific staining techniques. This result also demonstrates the usefulness of this methodology toward discovery of specific seed proteins that posses posttranslational modifications, that were previously unnoticed.

These same gels previously analyzed with ProQ Diamond, were then post-stained with SYPRO Ruby to compare both total protein yield and separation with the previous DIGE results. Ultimately, we wanted to ascertain how effective the method would be for obtaining gels conductive to protein spot isolation and eventual protein identification. The results were very similar to the previous DIGE analysis, showing a large reduction in the globulins and a massive enhancement of those non-abundant proteins

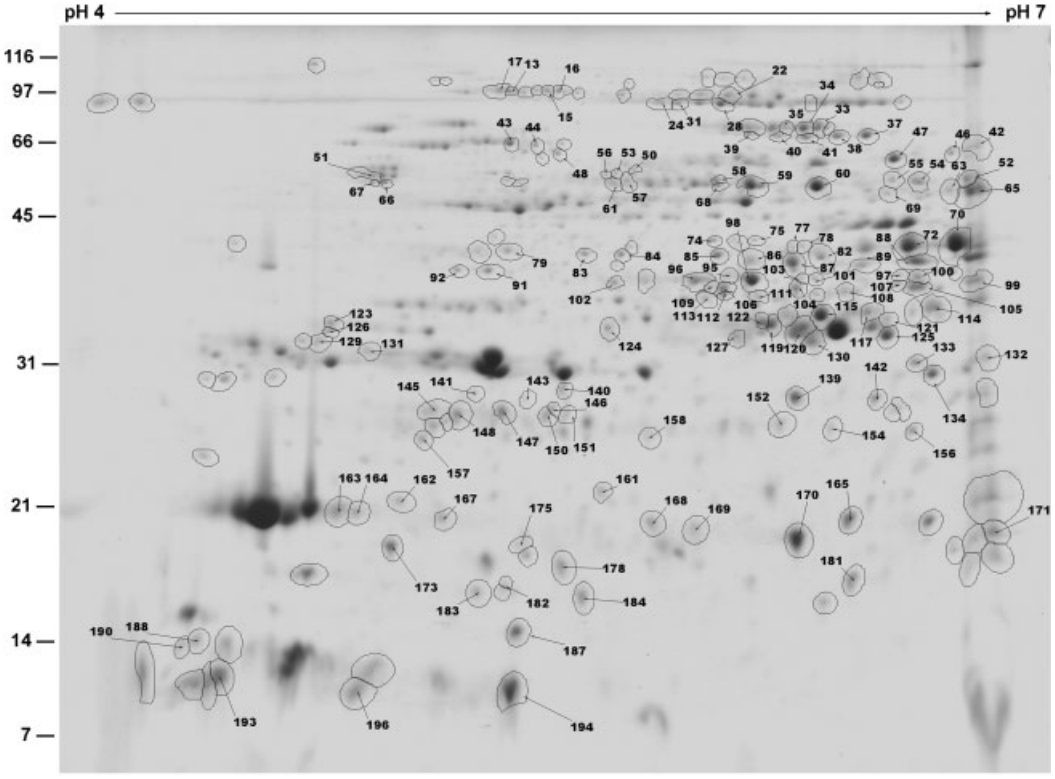

Figure 5. Image analysis using Delta2D software of the fractionated sample stained with Coomassie G-250 revealed a 2.5-fold increase in %volume in 197 out of 924 total spots. These now more apparent protein spots, detected using a relatively sensitive, visual and MS compatible stain such as Coomassie G-250, lend themselves more proficiently to isolation and analysis with peptide mass fingerprinting. All 197 spots were picked from this gel, processed and submitted for identification using MALDI-TOF MS as outlined in Section 2. Those protein spots identified are numbered and correspond to those proteins listed in Table 2. IEF separation was from pH 4 to 7 and protein molecular weight markers are in kDa.

Table 2. Proteins identified from soybean seed protein extract after using Ca^{2+} fractionation method, 2-DE separation and peptide mass fingerprinting (MALDI-TOF MS)

SID	Protein identification [species]	p*I* (Thr./Exp.)	M_r (Thr./Exp.)	MO	PM	SC	Accession	Database
13	Poly-[ADP-ribose] polymerase 3 (short) [*Glycine max*]	5.41/5.40	92 771/102 112	72	13	18	gil75213086	NCBInr
15	Expressed protein-like protein [*Glycine max*]	5.60/5.53	92 781/102 112	*223*	6	8	gil83853820	NCBInr
16	Poly-[ADP-ribose] polymerase 3 (short) [*Glycine max*]	5.41/5.57	92 771/102 112	107	14	23	gil75213086	NCBInr
17	Poly-[ADP-ribose] polymerase [*Glycine max*]	5.31/5.37	78 902/102 540	68	24	40	Glyma11g19070.1	MSDB
22	Elongation factor 2 [*Phaseolus vulgaris*]	5.93/6.12	94 708/99 113	59	10	13	EF2_BETVU	MSDB
24	Urease (embryo-specific) [*Glycine max*]	5.70/5.90	90 158/94 768	*1329*	8	14	Q7XAC5	UniProtKB
28	Urease (embryo-specific) [*Glycine max*]	5.68/6.10	90 841/93 640	100	16	25	gil32170829	NCBInr
31	Urease (embryo-specific) [*Glycine max*]	5.68/5.95	90 841/94 393	81	17	23	gil32170829	NCBInr
33	Methionine synthase [*Glycine max*]	5.93/6.41	84 401/79 357	116	20	33	gil33325957	NCBInr
34	Methionine synthase [*Glycine max*]	5.93/6.36	84 401/78 981	171	22	35	gil33325957	NCBInr
35	Methionine synthase [*Glycine max*]	5.93/6.30	84 401/79 357	97	10	17	gil33325957	NCBInr
37	Seed biotinylated protein of 65 kDa (BP75) [*Glycine max*]	6.10/6.57	67 894/74 470	155	24	41	gil75102139	NCBInr
38	Seed biotinylated protein of 65 kDa (BP75) [*Glycine max*]	6.10/6.47	67 894/74 093	116	17	30	gil75102139	NCBInr
39	Seed biotinylated protein of 65 kDa (BP75) [*Glycine max*]	6.10/6.18	67 897/75 597	$6.7e^7$	22	37	Q39846	UniProtKB
40	ABC transporter-like protein [*Glycine max*]	6.40/6.28	71 332/74 470	*210*	4	17	Q8W1S2	UniProtKB
41	Seed biotinylated protein of 65 kDa (BP75) [*Glycine max*]	6.10/6.37	67 894/73 718	105	15	30	gil75102139	NCBInr
42	β-Conglycinin (α-subunit) [*Glycine max*]	5.10/6.93	70 307/68 080	$5.0e^7$	18	28	Q94LX2	UniProtKB
43	Heat shock 70 kDa protein [*Glycine max*]	5.37/5.41	71 291/69 959	178	20	34	gil123601	NCBInr
44	Protease M3 thimet oligopeptidase (related) [*Glycine max*]	5.81/5.49	87 043/68 831	72	30	45	Glyma09g33490.1	MSDB
46	Cytochrome P450 97B2 [*Glycine max*]	6.50/6.84	64 742/65 122	*132*	5	12	O48921	UniProtKB
47	LEA-protein (related) [*Glycine max*]	6.33/6.65	50 644/63 325	$5.2e^5$	15	34	gil170010	NCBInr
48	Phosphoglucomutase (cytoplasmic) [*Glycine max*]	5.49/5.56	63 369/64 942	87	13	23	gil12585330	NCBInr
51	Protein disulfide isomerase-like protein [*Glycine max*]	5.06/4.96	58 963/58 834	131	16	33	gil49257109	NCBInr
52	51 kDa seed maturation protein [*Glycine max*]	6.65/6.89	51 065/57 756	93	12	23	gil414977	NCBInr
54	51 kDa seed maturation protein [*Glycine max*]	6.65/6.73	51 065/56 858	81	14	27	gil414977	NCBInr
55	Glucose-1-PO_4 adenylyltransferase (subunit 2) [*Vicia faba*]	6.19/6.65	56 310/57 397	59	5	16	gil1707944	NCBInr
58	LEA-protein (related) [*Glycine max*]	6.12/6.07	48 766/57 217	75	14	37	Glyma10g07410.1	MSDB
59	LEA-protein (related) [*Glycine max*]	6.12/6.19	48 766/55 600	78	23	42	Glyma10g07410.1	MSDB
60	LEA-protein (related) [*Glycine max*]	6.12/6.40	48 766/55 061	82	33	57	Glyma10g07410.1	MSDB
63	LEA-protein [*Glycine max*]	7.10/6.83	49 400/54 522	*2004*	10	16	gil311698	NCBInr
65	LEA-protein [*Glycine max*]	7.10/6.90	49 400/53 983	123	25	42	gil311698	NCBInr
66	Protein disulfide isomerase-like [*Glycine max*]	5.06/5.00	58 963/56 678	99	17	35	gil49257109	NCBInr
67	Protein disulfide isomerase-like [*Glycine max*]	5.06/4.96	58 963/56 678	70	11	19	gil49257109	NCBInr
68	LEA-protein [*Glycine max*]	6.12/6.08	48 766/55 959	59	19	42	Glyma10g07410.1	MSDB
69	Ferric leghemoglobin reductase-2 (precursor) [*Glycine max*]	6.90/6.63	53 003/53 624	*410*	8	23	O81413	UniProtKB
70	Alcohol dehydrogenase 1 [*Glycine max*]	6.20/6.85	42 007/42 687	*122*	5	13	gil22597178	NCBInr
72	Alcohol dehydrogenase 1 [*Glycine max*]	6.20/6.69	42 000/42 565	*122*	5	13	gil22597178	NCBInr
74	Chalcone synthase 2 [*Glycine max*]	6.40/6.06	42 505/43 235	*1825*	6	21	P17957	UniProtKB
75	MAP kinase 2 [*Glycine max*]	5.50/6.20	44 795/43 296	*589*	6	17	gil33340593	NCBInr
77	Pyruvate kinase [*Glycine max*]	6.80/6.32	54 361/42 687	*1468*	7	18	gil59668642	NCBInr
78	Glutamate dehydrogenase 2 [*Glycine max*]	6.10/6.35	44 687/42 745	131	5	15	Q5F2M8	UniProtKB
79	Actin-104 [*Glycine max*]	5.57/5.40	37 178/42 261	74	9	38	gil3219764	NCBInr
82	Phosphoglycerate kinase [*Pisum sativum*]	5.73/6.41	42 261/41 591	108	10	32	gil9230771	NCBInr
83	Ethylene responsive protein [*Glycine max*]	5.10/5.65	42 159/41 956	*973*	8	27	gil33331083	NCBInr
84	Nematode resistance HS1pro1 [*Glycine max*]	5.90/5.77	49 839/41 896	*1555*	5	15	gil194272792	NCBInr
85	Phosphoglycerate kinase [*Pisum sativum*]	5.73/6.08	42 261/41 713	100	10	30	gil9230771	NCBInr
86	Alcohol dehydrogenase [*Glycine max*]	6.10/6.18	36 381/41 266	*$1.6e^4$*	9	28	gil4039115	NCBInr

Table 2. Continued

SID	Protein identification [species]	p*I* (Thr./Exp.)	M_r (Thr./Exp.)	MO	PM	SC	Accession	Database
87	Alcohol dehydrogenase [*Glycine max*]	6.10/6.32	36 381/41 043	*1.3e*^{7}	15	36	gil4039115	NCBInr
88	Glyceraldehyde-3-PO_4 dehydrogenase [*Glycine max*]	6.72/6.72	36 855/41 165	77	9	34	gil85720768	NCBInr
89	Glyceraldehyde-3-PO_4 dehydrogenase [*Glycine max*]	6.72/6.54	36 855/40 861	88	10	36	gil85720768	NCBInr
91	Glutamine synthetase (cytosolic GS-1) [*Glycine max*]	5.50/5.34	38 845/40 252	*470*	6	23	gil121336	NCBInr
92	MAP kinase 2 [*Glycine max*]	5.50/5.23	44 795/40 374	*491*	5	20	gil33340593	NCBInr
95	Vestitone reductase [*Glycine max*]	5.70/6.11	35 860/39 826	*2167*	6	21	gil197215943	NCBInr
96	35 kDa seed maturation protein [*Glycine max*]	6.00/6.00	35 342/39 339	109	16	45	gil4102190	NCBInr
97	Pyruvate dehydrogenase kinase (mitochondrial isoform 1) [*Glycine max*]	7.30/6.66	41 705/39 826	*1410*	5	15	Q6PP99	UniProtKB
98	35 kDa seed maturation protein [*Glycine max*]	6.00/6.19	35 342/39 217	118	17	44	gil4102190	NCBInr
99	Gibberellin 3-β-hydroxylase [*Cucurbita maxima*]	6.90/6.90	40 379/39 339	*1483*	4	14	gil8247213	NCBInr
100	Glyceraldehyde-3-PO_4 dehydrogenase [*Glycine max*]	6.72/6.72	36 855/39 643	*1.7e*^{4}	7	29	gil85720768	NCBInr
101	35 kDa seed maturation protein [*Glycine max*]	5.96/6.39	35 320/39 522	86	14	40	gil4102190	NCBInr
102	Pyruvate kinase [*Glycine max*]	7.50/5.74	55 303/39 156	*1210*	9	21	Q42806	UniProtKB
104	Malate dehydrogenase (cytosolic) [*Glycine max*]	6.32/6.33	35 846/38 548	99	12	39	gil42521311	NCBInr
105	Seed maturation protein PM34 [*Glycine max*]	6.60/6.72	32 032/38 670	81	11	33	gil9622153	NCBInr
106	Malate dehydrogenase (cytosolic) [*Glycine max*]	6.32/6.09	35 846/38 487	121	12	35	gil42521311	MSDB
107	Catalase-1/2 [*Glycine max*]	6.80/6.65	56 847/38 974	*3019*	8	19	P29756	UniProtKB
108	Chalcone synthase 6 [*Glycine max*]	5.80/6.48	42 533/38 304	*5.6e*^{5}	9	21	P30080	UniProtKB
109	35 kDa seed maturation protein [*Glycine max*]	5.96/6.05	35 320/38 730	59	12	37	gil4102190	NCBInr
111	Seed maturation protein [*Glycine max*]	6.00/6.20	25 659/37 756	*7.7e*^{4}	9	37	Q42447	UniProtKB
112	Malate dehydrogenase (cytosolic) [*Glycine max*]	6.32/6.09	35 846/38 000	74	9	32	gil42521311	NCBInr
113	PEP carboxylase [*Glycine max*]	6.00/6.04	110 541/37 574	*4.9e*^{4}	26	30	gil45505269	NCBInr
114	Seed maturation protein PM34 [*Glycine max*]	6.60/6.77	32 032/36 661	85	10	41	gil9622153	NCBInr
115	Seed maturation protein [*Glycine max*]	6.02/6.40	25 644/35 809	143	19	56	gil170020	NCBInr
117	Seed maturation protein PM34 [*Glycine max*]	6.60/6.56	32 032/36 235	90	11	44	gil9622153	NCBInr
119	Dehydrin [*Glycine max*]	6.00/6.25	23 788/35 139	*8.7e*^{8}	15	50	gil37495451	NCBInr
120	Dehydrin [*Glycine max*]	6.07/6.33	23 704/34 591	138	17	62	gil497417	NCBInr
121	Dehydrin [*Glycine max*]	5.87/6.61	23 734/35 687	75	11	46	gil37495455	NCBInr
122	Dehydrin [*Glycine max*]	6.07/6.22	23 704/35 139	146	17	61	gil497417	NCBInr
123	Dehydrin [*Glycine max*]	5.87/4.83	23 734/35 139	*1.5e*^{8}	11	47	gil37495455	NCBInr
124	Dehydrin [*Glycine max*]	5.87/5.72	23 720/34 652	86	11	46	gil37495455	NCBInr
125	Short-chain dehydrogenase [*Glycine max*]	6.38/6.62	32 097/33 861	66	12	40	Glyma07g38790.1	MSDB
126	Chlorophyll a/b-binding protein [*Glycine max*]	5.10/4.82	27 847/34 409	*2842*	5	21	Q39831	UniProtKB
127	Dehydrin [*Glycine max*]	5.87/6.13	23 720/33 860	*753*	4	25	gil37495455	NCBInr
129	Uncharacterized protein [*Vitis vinifera*]	5.10/4.79	36 108/33 313	*283*	4	15	A5AI87	UniProtKB
130	Dehydrin [*Glycine max*]	5.90/6.37	23 734/33 435	*3211*	5	31	gil37495455	NCBInr
131	Soybean agglutinin, chain A complexed with 2,6-pentasaccharide [*Glycine max*]	5.15/4.95	27 555/32 461	59	7	39	gil6729836	NCBInr
132	Polyphosphoinositide binding protein Ssh1p [*Glycine max*]	6.80/6.93	36 939/31 791	*572*	5	18	O48939	UniProtKB
133	Putative RAN2 (small Ras GTP-binding nuclear protein) [*Allium sativum*]	6.65/6.71	25 274/31 243	124	12	52	gil87046111	NCBInr
134	Unknown protein [*Glycine max*]	6.56/6.76	27 277/30 401	80	15	79	Glyma11g36840.1	MSDB
139	Triosephosphate isomerase [*Glycine max*]	5.87/6.32	27 415/28 775	165	16	57	gil77540216	NCBInr
140	Fe-superoxide dismutase [*Glycine max*]	5.60/5.57	27 881/29 417	70	6	31	gil134646	NCBInr
141	Lectin (soybean agglutinin) [*Glycine max*]	5.70/5.29	30 928/29 074	*965*	4	21	gil126151	NCBInr
142	Unknown protein [*Glycine max*]	6.57/6.57	27 708/28 732	62	15	45	Glyma13g25870.1	MSDB
143	Fe-superoxide dismutase [*Glycine max*]	5.45/5.45	27 506/28 732	59	7	29	gil37654895	NCBInr
146	Glutathione *S*-transferase (GST) [*Glycine max*]	5.49/5.54	27 049/28 090	72	13	56	Glyma19g36080.1	MSDB
147	Seed maturation protein PM36 [*Glycine max*]	5.27/5.38	26 313/27 790	52	7	37	gil62287132	NCBInr
148	Seed maturation protein PM32 [*Glycine max*]	5.49/5.24	18 871/27 619	75	11	48	gil5733686	NCBInr
150	Ferritin-2 (chloroplastic) [*Glycine max*]	5.77/5.51	28 830/27 491	85	8	32	gil29839388	NCBInr
151	Ferritin-3 (chloroplastic) [*Glycine max*]	5.40/5.58	28 899/27 534	*220*	4	20	gil29839387	NCBInr

Table 2. Continued

SID	Protein identification [species]	p*I* (Thr./Exp.)	M_r (Thr./Exp.)	MO	PM	SC	Accession	Database
152	Glutathione *S*-transferase (GST 14) [*Glycine max*]	7.14/6.27	25 340/26 935	129	11	40	gil11385443	NCBInr
154	Chalcone isomerase A [*Glycine max*]	6.32/6.43	23 289/26 678	116	8	45	gil114199183	NCBInr
156	Manganese superoxide dismutase [*Glycine max*]	8.60/6.69	26 707/26 507	*999*	4	23	gil147945633	NCBInr
157	Transcription factor LEC1-A [*Glycine max*]	6.30/5.12	25 196/25 950	*198*	5	14	gil158525281	NCBInr
158	Glutathione *S*-transferase, GST [*Glycine max*]	5.98/5.85	23 491/26 164	70	9	62	Glyma10g43730.1	MSDB
161	Unnamed protein (similarity to ribonucleotide reductase) [*Vitis vinifera*]	5.50/5.69	23 963/22 313	*5201*	4	35	gil157336967	NCBInr
162	Trypsin inhibitor [*Glycine max*]	6.12/5.05	18 250/21 671	77	5	38	gil9367042	NCBInr
163	Trypsin inhibitor A OS [*Glycine max*]	4.99/4.85	24 275/21 129	60	6	24	ITRA_SOYBN	MSDB
164	Lactate/malate dehydrogenase α/β C-terminal domain containing protein [*Medicago truncatula*]	4.60/4.91	23 234/21 037	*3475*	4	25	gil124365520	NCBInr
165	Predicted protein [*Physcomitrella patens* subsp. *patens*]	7.00/6.48	21 174/20 697	*9487*	5	23	gil168003575	NCBInr
167	Glutathione *S*-transferase GST-13 [*Glycine max*]	5.50/5.19	25 198/20 759	*539*	4	19	Q9FQE5	UniProtKB
168	Unknown protein [*Medicago truncatula*]	5.71/5.85	17 494/20 512	83	6	40	gil217075034	NCBInr
169	Hsp22.3 (small heat-shock protein) [*Glycine max*]	5.90/5.98	22 317/20 234	*132*	4	23	Q39819	UniProtKB
170	Seed maturation protein PM31 [*Glycine max*]	6.10/6.31	17 907/19 679	91	9	46	gil4838149	NCBInr
171	Seed maturation protein PM30 [*Glycine max*]	8.95/6.94	15 145/20 049	*488*	6	26	gil4838147	NCBInr
173	Seed maturation protein PM22; LEA protein [*Glycine max*]	5.16/5.02	16 735/19 339	71	5	35	gil4585271	NCBInr
175	Predicted protein [*Physcomitrella patens* subsp. *patens*]	5.50/5.44	35 317/19 462	*572*	4	20	gil168012940	NCBInr
178	Phosphatidylinositol N-acetylglucosaminyl transferase (P Subunit) [*Glycine max*]	8.84/5.56	78 947/18 382	63	14	30	Glyma07g37150.1	MSDB
181	Nucleoside diphosphate kinase 1 [*Glycine max*]	5.93/6.48	16 489/17 703	78	5	30	gil2498078	NCBInr
182	Predicted protein [*Physcomitrella patens* subsp. *patens*]	5.60/5.38	16 496/17 332	*579*	4	31	A9RIS9	UniProtKB
183	Glycine-rich RNA-binding protein [*Glycine max*]	6.58/5.29	15 894/17 085	109	4	29	gil5726567	MSDB
187	LEA-protein [*Glycine max*]	5.52/5.42	11 485/15 110	73	6	43	gil1762955	NCBInr
188	Kunitz trypsin inhibitor (truncated) [*Glycine max*]	4.59/4.41	15 963/14 647	*281*	4	24	Q9ATY0	UniProtKB
190	Hypothetical protein [*Vitis vinifera*]	5.00/4.36	12 557/14 400	*4.9e*^{4}	4	35	gil147776031	NCBInr
193	Predicted protein [*Physcomitrella patens* subsp. *patens*]	4.50/4.48	13 950/12 276	*1340*	4	33	A9TGZ8	UniProtKB
194	Napin-type 2S albumin 1 precursor [*Glycine max*]	6.00/5.40	17 835/11 214	*329*	7	34	Q9ZNZ4	UniProtKB
196	Bowman-Birk proteinase inhibitor [*Glycine max*]	4.90/4.91	9468/11 256	*893*	4	40	P01064	UniProtKB

A large portion of the 197 spots enhanced more than 2.5-fold and visualized with Coomassie stain were isolated and identified. Spot identification numbers (SID) correspond to those protein spots labeled in Fig. 5. Isoelectric point (p*I*) and molecular weight (M_r) values are given as theoretical/experimental values. MOWSE scores (MO) in italics represent those searches performed *via* Protein Prospector using a peptide mass tolerance of 20 ppm. MASCOT ''probability-based MOWSE scores'' $\geq$57 represent those above 95% confidence limit ($p \leq 0.05$) using a peptide mass tolerance of 20 ppm. Proteins not meeting these criteria are not included in table. Peptides matched (PM), % sequence coverage (SC) and accession numbers within each respective database are given. Searches were confined to *Glycine max* databases unless no clear match could be found. Databases are as follows: National Center for Biotechnological Information (NCBI non-redundant); Matrix Sciences Database-MASCOT (MSDB); Universal Protein Knowledgebase (UniProtKB).

when calcium fractionation of the extracted proteins was used prior to separation. Those proteins removed from the seed extract with fractionation (Fig. 4A, numbered arrows) correspond to those proteins listed in Table 1. The non-fractionated soybean seed proteins when compared with the 10 mM $CaCl_2$ fractionated seed proteins (Fig. 4B) when stained with SYPRO Ruby clearly demonstrate the effectiveness of this methodology toward acquisition of protein spots for isolation and identification. Here, unlike the DIGE analysis, the non-fractionated and fractionated samples (equal protein load) were run on two separate gels to demonstrate the effectiveness of the method for future

protein isolation and identification with mass spectrometry techniques.

Identical results were obtained when additional, identically prepared non-fractionated and fractionated samples were prepared and two dimensionally separated using the same IEF range (pH 4–7) and a broader range (pH 3–10) and stained with colloidal Coomassie (gel comparison not shown). Dynamic resolution was higher however, when a narrower range first dimension separation was utilized in conjunction with SYPRO Ruby protein stain. SYPRO Ruby is a more sensitive stain than Coomassie G-250, and albeit not visually conductive to spot acquisition, still allows for isolation and identification of proteins after separation using specialized spot-picking instrumentation.

Generating intensity-based spot volumes (normalized) of each individual protein spot, Delta2D image analysis software revealed not only the massive amounts of α and α′ of β-conglycinin and basic subunit of glycinin in the non-fractionated sample, but the huge reduction of those proteins, along with reduction of the β subunit of β-conglycinin and acidic subunit of glycinin, in the Ca^{2+} fractionated sample. Percent spot volumes ratios [control]/[fractionated] and fold reduction data were generated for all individual calcium fractionations and 2-DE separations, using the three different visualization techniques (DIGE, SYPRO Ruby and Coomassie). However, spot volumes were compared within the same visualization technique, and fold reduction was compared across all image comparisons and is reported in Table 1. Overall, looking at several independent fractionation events and using several different visualization techniques, an 87% (±4%) reduction in globulin storage proteins is obtainable using 10 mM calcium fractionation of salt-soluble protein extract from soy seed.

Analysis of the SYPRO Ruby-stained gels revealed that 1184 total spots were detected in the calcium fractionated sample compared with only 891 in the non-fractionated seed extract. Delta2D analysis software revealed that 293 new spots were additionally detectable after fractionation and that there was a twofold increase in %volume with 541 spots out of the 1184 total spots detected. Within those 541 spots, 250 had ≥a threefold increase and 139 had ≥a fourfold increase. An analysis of those gels stained with Coomassie revealed a total of 924 spots detected using a 200 μg protein load, with an enhancement of 197 spots (≥2.5-fold %volume increase) when the seed extract was fractionated with 10 mM Ca^{2+} (Fig. 5). With 2.5 times the amount of protein, a large portion of those proteins were able to be isolated and analyzed with peptide mass fingerprinting *via* MALDI-TOF MS (Table 2). These data, along with analysis using fluorescent detection, clearly demonstrate not only the reproducibility, but the usefulness of this methodology toward our goal of enrichment of non-abundant protein spots for the analysis of global changes in the seed proteome and protein spot isolation and discovery.

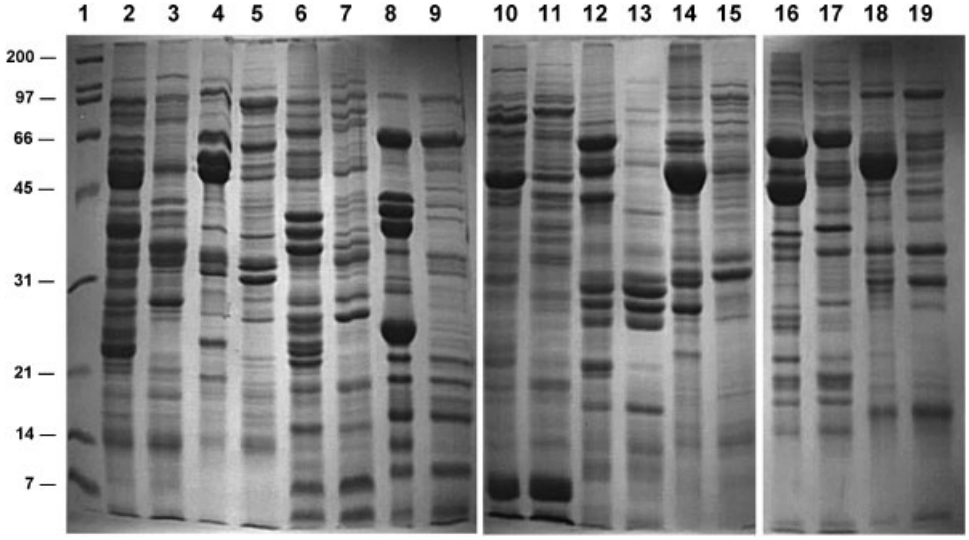

Figure 6. Fractionation of abundant seed proteins in various legumes with 10 mM Ca^{2+}. Salt- soluble proteins were extracted from various economically important legumes using the same low ionic strength buffer used with soybean. The effects of Ca^{2+} on protein fractionation were examined to determine the universality of the method for major seed protein removal in legumes. Lane 1, protein molecular weight markers in kDa; Lane 2, Smooth Vetch; Lane 3, Smooth Vetch (+Ca^{2+}); Lane 4, cowpea; Lane 5, cowpea (+Ca^{2+}); Lane 6, chick-pea; Lane 7, chick-pea (+Ca^{2+}); Lane 8, peanut; Lane 9, peanut (+Ca^{2+}); Lane 10, fenugreek; Lane 11, fenugreek (+Ca^{2+}); Lane 12, American potato; Lane 13, American potato (+Ca^{2+}); Lane 14, mung bean; Lane 15, mung bean (+Ca^{2+}); Lane 16, pigeon pea; Lane 17, pigeon pea (+Ca^{2+}); Lane 18, Adzuki bean; Lane 19, Adzuki bean (+Ca^{2+}). Calcium fractionation of seed storage proteins was not as pronounced in common bean, lima bush bean, black gram bean, garden Lupin, Bird's-foot Trefoil and alfalfa.

3.3 Application with other legumes

All legume seeds are particularly rich in globulin storage proteins. Homology among these proteins has been shown to be highly conserved throughout the legume family [10, 36]. Our successful reduction of glycinin and β-conglycinin in protein extracts from cultivated soybean seed brought us to hypothesize that the same or a similar technique might be universally applicable to other important crop seed legumes. Mature, dried seeds from peanut, alfalfa, various pea varieties, bean varieties, vetch, lupin, trefoil and American potato were ground into a powder and extracted using the same low ionic strength buffer that had been used with soybean. A second set of identical samples were fractionated using 10 mM calcium chloride, exactly as before. After clarification of the solution, samples of the supernatant were analyzed using 1-D SDS-PAGE. The results demonstrate the removal of several highly abundant proteins within many species of legume (Fig. 6) and the enhancement of those previously poorly resolved proteins. This suggests that this methodology is applicable to more than just soybean, and if used in conjunction with 2-D protein separation, could aid in the analysis and discovery of those less abundant seed proteins within each species. Interestingly, not all legume seed extracts were conductive to calcium fractionation. The affects of calcium addition to the seed extract of common bean, bush bean, gram bean, lupin, trefoil and alfalfa were not as pronounced (data not shown). Fractionation was evident, but removal of major seed storage proteins was not as prominent as with the others tested. While it is known that globulins among the legumes are highly conserved, apparently the action of calcium within the salt-soluble globulins is different and depends on specific protein characteristics and/or other protein species within their environment.

4 Concluding remarks

Soybean seed, from a proteome standpoint, consists of several thousand proteins. In our typical 2-DE separation of soybean seed, roughly 1000–1500 proteins can be detected on a gel image using standard protocols and fluorescent stains. Of those proteins, only the most abundant proteins (<500) can clearly be detected and/or identified, since the overwhelming majority (60–80%) of separated proteins is made up of storage proteins. From our viewpoint, not only do these storage proteins limit the amount of protein that one can load and separate using 2-DE, thereby limiting the accessibility of the less abundant proteins, but they greatly interfere with the electrophoretic separation of the other proteins in both dimensions as their quantity is increased.

We found that improvement of 2-DE resolution can be achieved through fractionation of the proteome using a customized fractionation technique. Cultivated soybean seed storage proteins consist primarily of two major storage protein complexes, glycinin and β-conglycinin; the globulins, which account for an overwhelming majority of the total seed protein. A simple, fast and inexpensive method to remove these storage proteins from soybean seed has been demonstrated here. With the simple addition of 10 mM calcium chloride to the salt-soluble soybean seed protein extract in low ionic strength buffer, the α, α′, and β subunits of β-conglycinin and the acidic and basic subunits of glycinin are nearly 90% reduced from the total seed protein extract. This methodology allowed for those non-abundant proteins in soybean seed to be more accessible for assessment of posttranslational modifications and ultimately, protein spot isolation and identification.

Here, we have demonstrated with cultivated soybean how such a fractionation technique has the potential to allow more clues about protein expression or modification for far more proteins than ever before. This will ultimately help many of those who struggle with analysis of legume nutritive potential, yield potential or environmental stress response on a protein expression level. Additionally, since legume globulin homology is known to be conserved, and we have demonstrated that this method could be applicable to many, while not all legumes, this methodology should help advance proteomic research of seeds in many legumes.

The authors are grateful to Mark Johnston, Sukla Lakshman and Sungchan Jang for their excellent technical assistance in data collection and analysis. The authors acknowledge the Charles W. Gherke Proteomics Center at the University of Missouri for use of their laser scanners for image acquisition and for the use of their image analysis software. Names are necessary to report factually on available data: however, the USDA neither guarantees nor warrants the standard of product, and the use of the name by USDA implies no approval of the product to the exclusion of others that may be suitable.

The authors have declared no conflict of interest.

5 References

[1] Mann, M., Pandey, A., Use of mass spectrometry-derived data to annotate nucleotide and protein sequence databases. *Trends Biochem. Sci.* 2001, *26*, 54–61.

[2] Herman, E. M., Helm, R. M., Jung, R., Kinney, A. J., Genetic modification removes an immunodominant allergen from soybean. *Plant Physiol.* 2003, *132*, 36–43.

[3] van Wijk, K. J., Challenges and prospects of plant proteomics. *Plant Physiol.* 2001, *126*, 501–508.

[4] Kersten, B., Burkle, L., Eckehard, J. K., Giavalisco, P. *et al.*, Large-scale plant proteomics. *Plant Mol. Biol.* 2002, *48*, 133–141.

[5] Roberts, J. K. M., Proteomics and a future generation of plant molecular biologists. *Plant Mol. Biol.* 2002, *48*, 143–154.

[6] Granier, F., Extraction of plant proteins for two-dimensional electrophoresis. *Electrophoresis* 1988, *9*, 712–718.

[7] Salekdeh, G. H., Komatsu, S., Crop proteomics: aim at sustainable agriculture of tomorrow. *Proteomics* 2007, *7*, 2976–2996.

[8] Koshiyama, I., in: Gottschalk, W., Muller, H. P. (Eds.), *Seed Proteins Biochemistry, Genetics, Nutritive Value (Advances in Agricultural Biotechnology)*, Kluwer Academic Publishers, Dordrecht 1983, pp. 427–450.

[9] Natarajan, S. S., Xu, C., Bae, H., Caperna, T. J., Garrett, W. M., Characterization of storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using proteomic analysis. *J. Agric. Food Chem.* 2006, *54*, 3114–3120.

[10] Ferreira, R. B., Franco, E., Teixeira, A. R., Calcium and magnesium-dependent aggregation of legume seed storage proteins. *J. Agric. Food Chem.* 1999, *47*, 3009–3015.

[11] Krishnan, H. B., Biochemistry and molecular biology of soybean seed storage proteins. *J. New Seeds* 2000, *2*, 1–25.

[12] Nielsen, N. C., in: Verma, D. P. S., Shoemaker, R. C. (Eds.), *Soybean: Genetics, Molecular Biology and Biotechnology*, CAB International, Wallingford, UK 1996, pp. 127–163.

[13] Thanh, V. H., Shibasaki, K., Major proteins of soybean seeds. A straight forward fractionation and their characterization. *J. Agric. Food Chem.* 1976, *24*, 1117–1121.

[14] Helm, R. M., Cockrell, G., Connaughton, C., Sampson, H. A. *et al.*, A soybean G2 glycinin allergen 1. Identification and characterization. *Int. Arch. Allergy Immunol.* 2000, *123*, 205–212.

[15] Ogawa, A., Samoto, M., Takahashi, K., Soybean allergens and hypoallergenic soybean products. *J. Nutr. Sci. Vitaminol.* 2000, *46*, 271–279.

[16] Sherwry, P. R., Napier, J. A., Tatham, A. S., Seed storage proteins: structure and biosynthesis. *Plant Cell* 1995, *7*, 945–956.

[17] Xiang, P., Beardslee, T. A., Zeece, M. G., Markwell, J., Sarath, G., Identification and analysis of a conserved immunoglobulin E binding epitope in soybean G1a and G2a and peanut Ara h 3 glycinins. *Arch. Biochem. Biophys.* 2002, *408*, 51–57.

[18] Natarajan, S., Xu, C., Bae, H., Bailey, B. A. *et al.*, Proteomic and genetic analysis of glycinin subunits of sixteen soybean genotypes. *Plant Physiol. Biochem.* 2007, *45*, 4366–4444.

[19] Nielsen, N. C., Dickinson, C. D., Cho, T. J., Thanh, V. H. *et al.*, Characterization of the glycinin gene family in soybean. *Plant Cell* 1989, *1*, 313–328.

[20] Renkema, J. M. S., Knabben, J. H. M., Vliet, T., Gel formation by β-conglycinin and glycinin and their mixtures. *Food Hydrocolloids* 2001, *15*, 407–414.

[21] Tierney, M. L., Bray, E. A., Allen, R. D., Ma, Y. *et al.*, Isolation and characterization of a genomic clone encoding the β-subunit of β-conglycinin. *Planta* 1987, *172*, 356–363.

[22] Cole, K. D., Cousin, S. L., Jr., Size exclusion chromatography of soybean proteins and isoflavones. *J. Agric. Food Chem.* 1994, *42*, 2713–2720.

[23] Miernyk, J. A., Johnston, M. L., Chemical cross-linking immobilized Concanavalin A for use in proteomic analyses. *Prepar. Biochem. Biotechnol.* 2006, *36*, 203–214.

[24] Hirotsuka, M., Terashima, M., Taniguchi, H., in: patent, U. S. (Ed.) 1988.

[25] Howard, P. A., Lehnhardt, W. F., Orthoefer, F. T., in: patent, U. S. (Ed.) 1983.

[26] Lehnhardt, W. F., Gibson, P. W., Orthoefer, F. T., in: patent, U. S. (Ed.) 1983.

[27] Riblett, A. L., Herald, T. J., Schmidt, K. A., Tilley, K. A., Characterization of beta-conglycinin and glycinin soy protein fractions from for selected soybean genotypes. *J. Agric. Food Chem.* 2001, *49*, 4983–4989.

[28] Thiering, R., Hofland, G., Foster, N., Witkamp, G.-J., Van de Wielen, L., Fractionation of soybean proteins with pressurized carbon dioxide as a volatile electrolyte. *Biotechnol. Bioeng.* 2001, *73*, 1–11.

[29] Wu, S., Murphy, P. A., Johnson, L. A., Fratzke, A. R., Reuber, M. A., Pilot-plant fractionation of soybean glycinin and β-conglycinin. *J. Am. Oil Chem. Soc.* 1999, *76*, 285–293.

[30] Yamauchi, F., Sato, M., Sato, W., Kamata, Y., Shibasaki, K., Isolation and identification of a new type of β-conglycinin in soybean globulins. *Agric. Biol. Chem.* 1981, *45*, 2863–2868.

[31] Sakakibara, K., Noguchi, H., Interaction of 11S fraction of soybean protein with calcium ion. *J. Agric. Biol. Chem.* 1977, *41*, 1575–1580.

[32] Rao, A. G. A., Rao, M. S. N., Binding of Ca (II), Mg (II), and Zn (II) by 7S fraction of soybean proteins. *J. Agric. Food Chem.* 1976, *24*, 490–494.

[33] Kroll, R. D., Effect of pH on the binding of calcium ions by soybean proteins. *Cereal Chem.* 1984, *61*, 490–495.

[34] Yuan, Y. J., Velev, O. D., Chen, K., Campbell, B. E. *et al.*, Effect of pH and Ca^{2+}-induced associations of soybean proteins. *J. Agric. Food Chem.* 2002, *50*, 4953–4958.

[35] Deak, N. A., Murphy, P. A., Johnson, L. A., Fractionating soybean storage proteins using Ca^{2+} and $NaHSO_3$. *J. Food Sci.* 2006, *71*, C413–C424.

[36] Wang, T. L., Domoney, C., Hedley, C. L., Casey, R., Grusak, M. A., Can we improve the nutritional quality of legume seeds? *Plant Physiol.* 2003, *131*, 886–891.

[37] Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970, *227*, 680–685.

[38] Bradford, M. M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* 1976, *72*, 248–254.

[39] Sarma, A. D., Oehrle, N. W., Emerich, D. W., Plant protein isolation and stabilization for enhanced resolution of two-dimensional polyacrylamide gel electrophoresis. *Anal. Biochem.* 2008, *379*, 192–195.

[40] Agrawal, G. K., Thelen, J. J., Development of a simplified, economical polyacrylamide gel staining protocol for phosphoproteins. *Proteomics* 2005, *5*, 4684–4688.

