

Protein Profile of Mature Soybean Seeds and Prepared Soybean Milk

Anna Laura Capriotti, Giuseppe Caruso, Chiara Cavaliere,* Roberto Samperi, Serena Stampachiacchiere, Riccardo Zenezini Chiozzi, and Aldo Laganà

Dipartimento di Chimica, Università di Roma "La Sapienza", Piazzale Aldo Moro 5, 00185 Rome, Italy

S Supporting Information

ABSTRACT: The soybean (*Glycine max* (L.) Merrill) is economically the most important bean in the world, providing a wide range of vegetable proteins. Soybean milk is a colloidal solution obtained as water extract from swelled and ground soybean seeds. Soybean proteins represent about 35–40% on a dry weight basis and they are receiving increasing attention with respect to their health effects. However, the soybean is a well-recognized allergenic food, and therefore, it is urgent to define its protein components responsible for the allergenicity in order to develop hypoallergenic soybean products for sensitive people. The main aim of this work was the characterization of seed and milk soybean proteome and their comparison in terms of protein content and specific proteins. Using a shotgun proteomics approach, 243 nonredundant proteins were identified in mature soybean seeds.

KEYWORDS: soybean, soy milk, *Glycine max*, proteome, seed storage proteins, liquid chromatography–mass spectrometry

INTRODUCTION

The soybean (*Glycine max* (L.) Merrill) is an herbaceous plant of the leguminous family, originally from East Asia, and cultivated for food. Economically, the soybean is the most important bean in the world, providing vegetable proteins with high biological value¹ for millions of people and ingredients for hundreds of chemical products. Soybean seeds are very rich in nutritive components and represent an important source of vitamins and minerals—rich in unsaturated fatty acids (mono- and polyunsaturated) but poor in saturated fatty acids. The popularity of the soybean is also due to its high content of isoflavones, even if their effect on human health is quite controversial.²

Soybean proteins represent about 35–40% on a dry weight basis of which 70–83% have storage functions.³ Because soybean proteins contain all the amino acids that are essential to human nutrition, they are a great substitute for animal proteins.

Recently, due to its known health benefits,^{4,5} the use of soybean in human nutrition has significantly increased. Besides soybean seeds, several soybean products are available on the market, such as milk, sprouts, sauce, tofu, miso, flour, among others; moreover, due to their excellent nutritional and functional properties (e.g., water and fat absorption, emulsification, foaming, gelatinization, and binder), soybean proteins are widely used as ingredients in formulated foods, meat and poultry products, bakery and pastry products, dairy products, and others. However, soybean proteins can also be a source of allergens;^{6,7} at least 16 potential soy protein allergens have been identified;⁸ however, their relative clinical significance is still unknown.

Among soybean-based products, soy milk is one of the most consumed. It is essentially a colloidal solution obtained as water extract from swelled and ground soybeans; therefore, almost all the components (proteins, lipids, and saccharides) of the soybean seeds are present in the derived milk.⁹

Despite the importance and the widespread of soybeans and soy milk, nowadays, only a few studies are devoted to the characterization of the soybean seed proteome, and often the interest is focused on a limited number of proteins;^{10–16} furthermore, to our knowledge, no studies have been performed on the proteome of soybean milk.

To compensate for this lack and to evaluate possible protein differences between soybean seeds and milk, the aim of this work was to characterize the protein profile of these two sample typologies. A shotgun proteomic approach was used in order to provide a deeper characterization of the two protein profiles and highlight possible differences.

MATERIALS AND METHODS

Chemicals and Standards. All chemicals, reagents, and organic solvents of the highest grade available were purchased from Sigma-Aldrich (St. Luis, MO, U.S.A.) unless otherwise stated. The sequencing-grade modified trypsin and protease inhibitor cocktail were from Promega (Madison, WI, U.S.A.). Ultrapure water (resistivity 18.2 MΩ cm) was obtained by an Arium water purification system (Sartorius, Göttingen, Germany).

Solid phase extraction (SPE) C18 cartridges were BOND ELUT (Varian, Palo Alto, CA, U.S.A.).

Plant Materials. The mature soybean seeds and soybean milk employed in this study were kindly donated by Consorzio Agrario dell'Emilia (Società Cooperativa, San Giorgio di Piano, BO, Italy). The cultivar of sample was guaranteed by the suppliers and was cv. PR91M10.

Preparation of Soybean Milk. Soybean milk was prepared according to the Chinese method. Briefly, mature soybean seeds were thoroughly washed and soaked overnight at room

Received: July 18, 2014

Revised: September 16, 2014

Accepted: September 17, 2014

Published: September 17, 2014

temperature (RT) with 10 times their weight of distilled water. Using a kitchen mixer, the soybean seeds were homogenized at low speed for 1 min. Soybean milk was obtained from the resulting slurry by the removal of an insoluble residue (soy pulp fiber) by filtration (FILTER-LAB qualitative filter paper, pore size 43–48 μm). The soybean milk was heated in boiling water for 15 min and then cooled to RT.

Protein Extraction. The three protocols used to prepare the protein extracts, one for soybean seed samples and two for the soybean milk samples, are outlined in Figure 1.

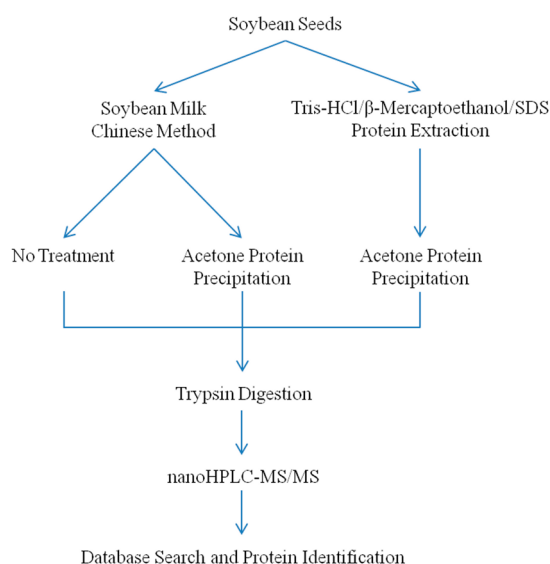


Figure 1. Workflow diagram for the experimental procedure.

Extraction of Proteins from Soybean Seeds. Soybean seeds were shortly sterilized with 70% ethanol for 10 s and then washed with distilled water three times. Dry seeds (ca. 1 g) were frozen in liquid nitrogen and ground to a fine powder in a precooled mortar with pestle. The powder (0.1 g) was defatted three times with 1 mL of petroleum ether, shaking gently and thoroughly for 15 min. Then, the proteins were extracted with 1 mL of a solution containing 50 mmol L⁻¹ Tris-HCl pH 8.8, 1.5 mmol L⁻¹ KCl, 0.07% β -mercaptoethanol, 1% protease inhibitor cocktail, and 1% (w/v) sodium dodecyl sulfate (SDS). The samples were incubated on ice for 1 h with intermittent mixing with vortex (1 min) every 15 min, and insoluble materials were removed by centrifugation at 4 °C for 15 min at 11 000g. The supernatant was transferred to a new centrifuge tube, four volumes of cold acetone with 0.07% (v/v) β -mercaptoethanol were added, and then the sample was mixed thoroughly and incubated at -20 °C overnight. The precipitate was collected by centrifugation (18 400g for 15 min at 4 °C), washed three times with acetone containing 0.07% (v/v) β -mercaptoethanol, air-dried, and solubilized in urea buffer pH 8.5 (8 mol L⁻¹ urea in 50 mmol L⁻¹ NH₄HCO₃). The proteins were quantified by Bradford assay using BSA standard¹⁷ and stored at -80 °C until the trypsin digestion. Three experimental replicates were performed.

Proteins from Soybean Milk without Any Treatment. From a 10 mL aliquot of soybean milk, prepared as described above, the proteins were directly quantified by Bradford assay using BSA standard¹⁷ and stored at -80 °C until further processing.

Acetone Precipitation of Proteins from Soybean Milk. The soybean milk (10 mL) was placed in acetone-compatible tube,

and four volumes of cold (-20 °C) acetone with 0.07% (v/v) β -mercaptoethanol were added. The tube was vortex-mixed and incubated overnight at -20 °C. The precipitated proteins from soybean milk were collected by centrifugation (18 400g for 15 min at 4 °C), washed three times with acetone containing 0.07% (v/v) β -mercaptoethanol, air-dried, and solubilized in urea buffer pH 8.5 (8 mol L⁻¹ urea in 50 mmol L⁻¹ NH₄HCO₃). The protein content was quantified by Bradford assay using BSA standard¹⁷ and stored at -80 °C until the trypsin digestion. Three experimental replicates were performed.

Enzymatic Digestion and Off-Line Desalting. Aliquots from soybean seed samples, soybean milk samples, and precipitated proteins from soybean milk samples were reduced, alkylated, and digested with trypsin. For the digestion of the proteins from soybean milk without any treatment, we added urea to the soybean milk to reach a 8 mol L⁻¹ concentration in order to denature the proteins. Reduction of disulfide bonds was performed with dithiothreitol (DTT, 200 mmol L⁻¹) in incubation at 37 °C for 1 h, under slight agitation. Carbamidomethylation of thiol groups was performed by addition of iodoacetamide (IAA, 200 mmol L⁻¹) and incubation for 1 h in the dark at RT. To consume any leftover alkylating agent and to avoid trypsin alkylation, DTT (200 mmol L⁻¹) was added, and samples were incubated at 37 °C for 1 h, under slight agitation. The samples were then diluted with NH₄HCO₃, pH 8.5 (50 mmol L⁻¹) to obtain a final 1 mol L⁻¹ urea concentration. Sequencing-grade modified trypsin was added (1:20, w/w, enzyme to protein ratio), and the samples were incubated overnight at 37 °C. Enzymatic digestion was quenched with trifluoroacetic acid (TFA). Digested samples were desalted using SPE C18 cartridges conditioned with ACN and rinsed with 0.1% TFA aqueous solution. Peptides were eluted from the SPE column with ACN/ddH₂O (50/50, v/v) containing 0.05% TFA, and were dried in a Speed-Vac SC 250 Express (Thermo S 164 avant, Holbrook, NY, U.S.A.). Each sample was reconstituted with 0.1% HCOOH aqueous solution and stored at -80 °C until analysis.

NanoHPLC-MS/MS Analysis. Nano high-performance liquid chromatography (nanoHPLC) coupled to tandem mass spectrometry (MS/MS) analysis was performed on a hybrid linear ion trap-Orbitrap mass spectrometer (model Orbitrap Elite, Thermo Scientific, Bremen, Germany) equipped with a nanoelectrospray ion source. Peptide mixtures were separated by reversed-phase (RP) chromatography using the Dionex Ultimate 3000 (Dionex Corporation Sunnyvale, CA, U.S.A.). The LC system was connected to a manufactured 25 cm fused-silica nanocolumn, 75 μm i.d., packed in-house with Acclaim-C18 2.2 μm silica microparticles, with outlet frit prepared using Kasil. Peptide mixtures were enriched on a 300 μm i.d. $\times$ 5 mm Acclaim PepMap 100 C18 (5 μm particle size, 100 Å pore size) precolumn (Dionex), employing a premixed mobile phase H₂O/ACN 98:2 (v/v) (from loading pump) containing 0.1% (v/v) HCOOH at a flow-rate of 10 $\mu\text{L min}^{-1}$. LC gradient was optimized to detect the largest set of peptides, using H₂O/HCOOH (99.9:0.1, v/v) as mobile phase A and ACN/HCOOH (99.9:0.1, v/v) as mobile phase B. After an isocratic step at 1% B for 5 min, B was linearly increased to 11% within 2 min and then to 31% within 120 min; afterward, phase B was maintained at 31% within 10 min, and increased to 80% within the following 10 min. Then, phase B was maintained at 80% for 10 min to rinse the column. Finally, B was lowered to 1% over 1 min and the column re-equilibrated for 22 min (180 min total

run time). MS spectra of tryptic digests were collected over an m/z range of 380–2000, using a resolution setting of 60 000 (full width at half-maximum, fwhm, at m/z 400), operating in the data-dependent mode to automatically switch between Orbitrap-MS and linear ion trap-MS/MS acquisition. MS/MS spectra were collected for the 20 most abundant ions in each MS scan. Rejection of +1, and unassigned charge states was enabled. All MS/MS spectra were collected using normalized collision energy of 30%, and an isolation window of 2 m/z . Ion trap and Orbitrap maximum ion injection times were set to 100 and 200 ms, respectively. Automatic gain control (AGC) was used to prevent overfilling of the ion traps and was set to 1×10^6 for full FTMS scan, and 1×10^4 ions in MSⁿ mode for the linear ion trap. To minimize redundant spectral acquisitions, dynamic exclusion was enabled with a repeat count of 1 and a repeat duration of 30 s with exclusion duration of 70 s. In order to increase the number of identified peptides/proteins, we performed three technical replicates (nanoHPLC-MS/MS runs) for each of the three experimental replicates.

Database Searching and Protein Identification. Raw MS/MS data files from Xcalibur software (version 2.2 SP1.48, Thermo Fisher Scientific) were submitted to Proteome Discoverer software (version 1.3, Thermo Scientific) with the Mascot (v.2.3.2, Matrix Science) search engine for peptide/protein identification. The searches were performed against Swiss-Prot database (Release 01-2014, 541 954 sequences). The search was limited to proteins from species of the *Other Green Plants* (18 092 sequences) taxonomy entries and performed using the built-in decoy search option of Mascot. Enzymatic digestion with trypsin was selected, together with maximum of 2 missed cleavages, peptide charges from +2 to +5, a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.8 Da; acetylation (N-term), oxidation (M), and deamidation (N, Q) were used as dynamic modifications, whereas carbamidomethylation (C) was used as static modification.

Scaffold Analysis. Scaffold software (version Scaffold 3.4.3, Proteome Software Inc., Portland, OR, U.S.A.)¹⁸ was used to validate MS/MS based peptide and protein identification. The additional X! Tandem search engine (The GPM, Cyclone version 2010.12.01.1) was also used, keeping the same parameters previously used for Mascot. According to the Peptide and Protein Prophet algorithms^{19,20} implemented into Scaffold, identifications were accepted if they could be established at a probability greater than 95% and 99% for proteins and peptides, respectively, and if proteins contained at least 2 unique peptides, resulting in a false discovery rate of 0.02% for peptides and 2.1% for proteins. Proteins that contained similar peptides and could not be differentiated on the basis of MS/MS analysis alone were grouped to satisfy the principles of parsimony.

RESULTS AND DISCUSSION

Design of the Experiment and Proteomic Approach.

The objective of this study was to characterize the protein profile of mature seeds of soybean and soybean milk obtained from the seeds themselves using the Chinese method. Furthermore, precipitation with acetone was also evaluated in order to assess whether the protein precipitation of soybean milk could improve the subsequent nanoHPLC-MS/MS analysis and the identification of proteins. Therefore, a shotgun proteomic approach was employed. The proteins extracted from the mature soybean seeds, the proteins of the soybean

milk and those obtained from soybean milk by precipitation with acetone were digested with trypsin, and the resulting peptides were characterized using nanoHPLC-MS/MS analysis and database search. Figure 1 shows a workflow diagram for the experimental procedures.

Proteins Identified in Mature Soybean Seeds and Soybean Milk. This is the first study reporting the comparison of protein profiles of mature soybean seed and soybean milk by a shotgun proteomic approach. The analysis of the protein profile of the three systems under investigation led to the identification of 209 proteins in seeds, 200 in milk, and 183 in precipitated soybean milk. Figure 2 shows a Venn diagram in

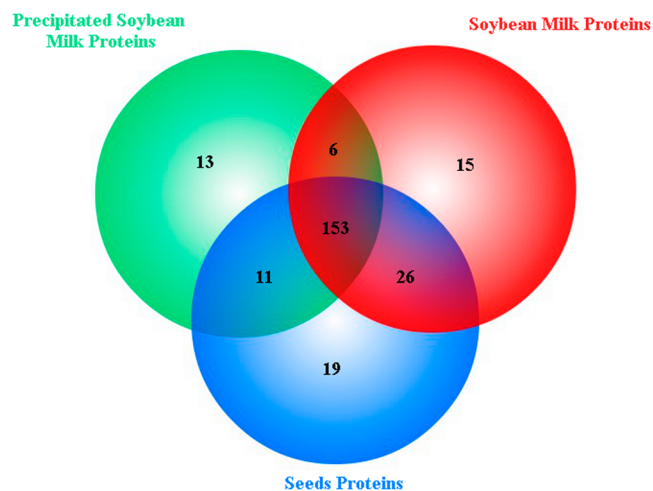


Figure 2. Venn diagram of the differentially identified proteins in soybean seeds, soybean milk, and precipitated soybean milk.

which the protein profiles of the three systems investigated are compared. From a first analysis, the protein profiles of the three systems apparently showed some differences. The Venn diagram shows 19 unique proteins for the seeds, 15 unique proteins for the soybean milk, and 13 unique proteins for the precipitated soybean milk. These differences are, however, misleading. The milk samples, and in particular the precipitated milk samples, cannot present unique proteins, because they derive from the same seeds, the proteome of which was also characterized (see Supporting Information file). Thus, the unique proteins identified for the milk and the precipitated milk samples are to be ascribed to the three different approaches for obtaining the protein samples and to the statistical variability of the LC-MS/MS experiments.

Proteins extracted from the mature soybean seeds were obtained with a standard extraction procedure, which requires the use of specific reagents (Tris-HCl, KCl, β -mercaptoethanol, protease inhibitor cocktail, and SDS) and a precipitation step with acetone (Figure 1); on the contrary, the soybean milk proteins were obtained with a simple but effective extraction from the aqueous phase at RT (Figure 1). In our opinion, the two protein extraction methods provide comparable results; the few differences exhibited by protein profiles (unique proteins) are due to the statistical variability of the experiments. Indeed, small changes in the signal intensity of a peptide (especially if present at low concentration) affect its selection for MS/MS spectrum acquisition, which can therefore be different between different LC-MS experiments. The quality of the MS/MS experiments can also vary, leading to successful identification in some cases but not in others.²¹ Therefore, this shotgun

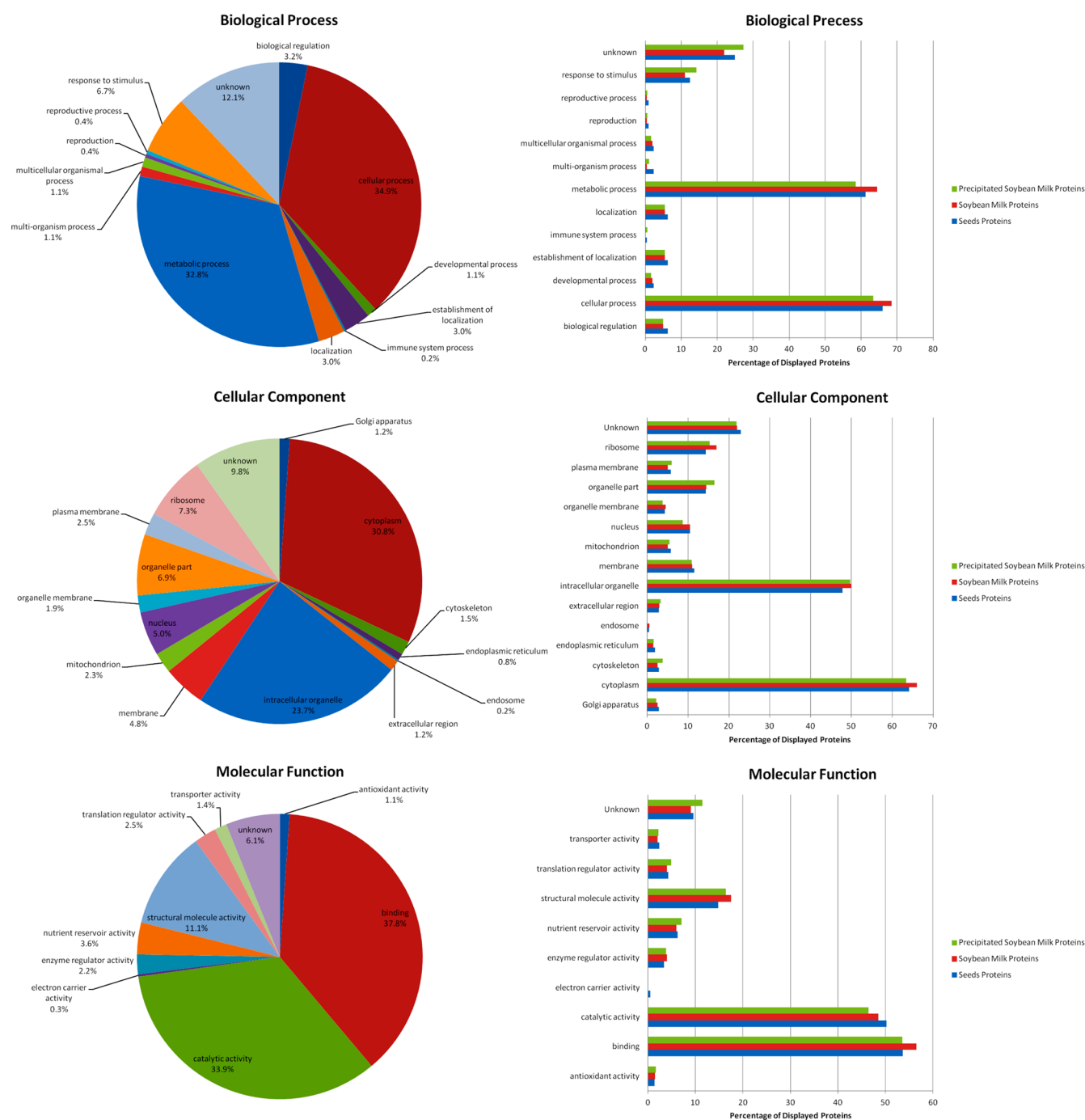


Figure 3. Gene Ontology annotation according to biological process, cellular component, and molecular function of the 243 nonredundant identified proteins and relative distribution of the seed proteins, soybean proteins, and precipitated soybean milk proteins.

proteomics analysis showed substantial equivalence of the soluble protein composition in mature soybean seeds and prepared soybean milk.

The largest differences presented by the third system (precipitated soybean milk) compared to the other two (Figure 2) are mainly due to the different procedure by which the proteins were obtained (i.e., the precipitation from soybean milk with acetone).

Precipitation is frequently used to concentrate proteins and to remove interfering compounds which could be responsible for irreproducible results. This is the case, for example, of most agriculturally interesting species, where the precipitation step is

necessary to remove secondary metabolite compounds, which are present in large amounts. The precipitation with acetone or, even more significantly, the precipitation with acetone/TCA definitely improve the results obtained by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) analysis because it allows a better resolution.¹⁴ However, in the shotgun proteomics approach, the precipitation leads to a loss of information, i.e. a smaller number of identified proteins (Figure 2). Our study showed that the prepared soybean milk, which contains sugars, lipids, and secondary metabolic compounds over proteins, can be analyzed by the shotgun proteomics approach directly, without any precipitation step with acetone.

The protein profile of the mature soybean seeds provided in this study is the most comprehensive one to the authors' best knowledge. Ten years ago, two studies based on 2D-PAGE, which is an extremely time-consuming technique, were published on the proteomic analysis of mature soybean seeds; Herman et al.¹¹ identified 111 protein spots, and Mooney et al.¹³ identified 44 protein spots. A more recent study, based on nanoUPLC-MS^E technique,²² identified 113 proteins; Gomes et al.¹⁵ identified 117 proteins with one-dimensional gel electrophoresis followed by MS and MS/MS analysis. Our results far exceed the low number of proteins identified in these previous works and provide additional identifications for the proteins present at low amounts in addition to those particularly abundant, such as the seed storage proteins. However, recently, Miernyck et al.²³ reported the identification of proteins in different stages of maturation only in the testa from developing soybean seeds. In almost every maturation stage Miernyck et al.²³ identified a larger number of proteins than the one obtained from the proteomic analysis of the whole mature soybean seed. This difference can be explained by several reasons: (i) In the mature whole seeds, 70–80% of the protein content is made up of storage proteins, which are highly abundant and thus hinder the detection of the lower abundance proteins, (ii) although in this work, a specific database for *Glycine max* was not employed; SwissProt was used, instead, because it is a high quality, curated, annotated, and non-redundant protein sequence database. Despite this, our results (243 identified proteins) are not excessively lower than the ones reported by Miernyck et al.²³ for the last stages of development of the testa (193 identified proteins in S8 and 272 identified proteins in S9).

This study demonstrated that a shotgun proteomics approach with in solution digestion is not sensitive to the presence of lipids, saccharides, and secondary metabolites that are in the soybean milk. Vice versa, a 2D-PAGE analysis would have been sensitive to the presence of contaminants of any nature and therefore requires a precipitation step that improves the resolution and reproducibility of the gels but leads to loss of information, with fewer identified proteins.

Protein Classification. In the present study, a total of 243 nonredundant proteins was identified (Figure 2). However, 161 proteins were annotated with a taxonomy other than *Glycine max*, like *Nicotiana tabacum*, *Pisum sativum*, or *Zea mays*. This could be ascribed to the fact that soybean is a newly sequenced species and several proteins are still not annotated in SwissProt. Consequently, we blasted all the proteins in the entire UniProtKB database, filtering the results by *Glycine max* only. All the blast searches gave good results (the majority with more than 85% of identity and maximal E-values of 10^{-18}) with proteins of soybean present only in TrEMBL section (see Supporting Information, Tab: BLAST).

The identified proteins were classified by the Gene Ontology (GO) terms in three broader domains: biological process, cellular component, and molecular function, and according to the Scaffold's system, each protein displayed in the samples may have one or many GO terms (Figure 3). With regard to the molecular function domain, the two most represented categories are binding and catalytic activity, with a large number of proteins belonging to both categories. This finding is in agreement with the classification based on the biological process, where the two most represented categories are cellular process and metabolic process, and also in this case many proteins belong to both categories. These proteins are often

referred to as “housekeeping” proteins, which are essential for the maintenance of normal cell metabolism and involved in seed development, and some of these proteins are enzymes or enzyme inhibitors. The housekeeping protein fraction is made up of relatively small amounts of numerous protein species.

The other important category of proteins in molecular function domain is nutrient reservoir activity (Figure 3), better known as storage proteins. The two major seed storage proteins from soybean are multimeric with Svedberg coefficients of 7S and 11S. The 7S globulins are composed of β -conglycinin subunits, whereas 11S globulins are composed of glycinin proteins.^{24–26} The storage proteins account for the major part of the seed proteins, 70–83% of the total proteins in terms of abundance, and are represented by relatively few different species of proteins. The main characteristics of seed storage proteins are that (i) they accumulate in seeds in large quantities, (ii) they are tissue-specific and occur only in the seed, (iii) they are synthesized during seed development, (iv) they are often sequestered in membrane-bound organelles (protein bodies), and (v) they are hydrolyzed to release their constituent amino acids during germination.

There are several housekeeping proteins, but they are present in a small amount; vice versa, the storage proteins comprise only few proteins, but they are present in a large amount, as confirmed by MS/MS data (Supporting Information). Indeed, the storage proteins showed high values of spectral counts, whereas the housekeeping proteins showed low values of spectral counts.

Liu et al.²⁷ demonstrated a linear relationship between the spectral count and the relative protein abundance over a dynamic range of 2 orders of magnitude. The spectral counting method relies on the simple concept that an increase in protein abundance can increase the number of its detectable proteolytic peptides (i.e., the total number of MS/MS spectra that are acquired for that peptide).

With regard to domain cellular component, the two most represented categories are cytoplasm and intracellular organelle. This result confirms that most of the seed proteins are storage proteins. Storage proteins primarily accumulate in the protein storage vacuoles of terminally differentiated cells of the embryo and endosperm and as protein bodies (PBs) directly assembled within the endoplasmic reticulum.

PBs form as a consequence of developmentally regulated events that induce storage protein synthesis in specialized cells and promote storage protein accumulation in specific organelles. Storage proteins may remain in the endoplasmic reticulum or be transported through the endomembrane system to distal sites. The major seed vacuolar storage proteins, (i.e., 7S and 11S globulins) form dimers, trimers, and tetramers in the endoplasmic reticulum lumen shortly after synthesis.

Figure 3 shows that the relative distribution of the seed proteins, soybean proteins and precipitated soybean milk proteins, on the basis of GO classification, is substantially equivalent.

Identified Allergens in Soybean. The soybean, a legume of high nutritional value, is also known to be one of the major allergenic foods. Identification and characterization of specific protein components linked to soybean allergy has been the focus of several studies. The results of these investigations have shown that a heterogeneous group of proteins are IgE-binding and are potential allergens. In the present study, some of the proteins already described as allergens were found (Table 1).

Table 1. Identified Allergens in Soybean

no.	identified proteins	entry name
1	hydrophobic seed protein	HPSE_SOYBN
2	profilin 1 ^a	PROF1_SOYBN
3	profilin 2 ^a	PROF2_SOYBN
4	stress-induced protein SAM22	SAM22_SOYBN
5	beta-conglycinin, alpha chain	GLCA_SOYBN
6	34 kDa maturing seed protein	P34_SOYBN
7	glycinin	GLYG5_SOYBN
8	2S albumin	2SS_SOYBN
9	lectin (agglutinin)	LEC_SOYBN
10	trypsin inhibitor A	ITRA_SOYBN

^aProfilin was found with only one unique peptide and thus is not reported in the Supporting Information file.

The hydrophobic seed proteins are a group of proteins with low molecular weight which belong to the lipid transfer proteins (LTP) family,²⁸ the role of which is to transport the phospholipids from liposomes to mitochondria. It has been observed that these proteins are responsible for soybean epidemic asthma,²⁹ but they are not responsible for asthma that affects the workers of intermediate products of the soybean, which is instead due to proteins at higher molecular weight.³⁰

Profilin 1 and profilin 2 are proteins involved in the growth process of actin microfilaments and belong to a group of structural and metabolic proteins. The sequences that are capable of binding to IgE are maintained in all the proteins belonging to the family of profilin. They are the main cause of the allergic response to pollen. Rihs et al.³¹ have shown that the ability of profilin to bind IgE depends on the structural and conformational integrity.

Stress-induced protein SAM22 belongs to the BetVI family. It is involved in defense response of the plant. This protein is the leading cause of allergy to soybean in patients with birch pollinosis. Mittag et al.,³² through a clinical investigation, revealed that 96% of patients reacting to *Bet v 1* have also *Glym* 4-specific IgE and that 64% of them recognized other soybean proteins.

Beta-conglycinin is a seed storage glycoprotein which is accumulated during the seed growth and stored in order to be later hydrolyzed and used as a source of carbon and nitrogen atoms for the plant. The alpha-subunit of beta-conglycinin is responsible for the allergic response.³³

The 34 kDa maturing seed protein, also known as P34, is considered a major allergenic protein of soybean seeds.³⁴ It belongs to the peptidase C1 family and is involved in the hydrolysis of peptide bonds by a mechanism in which the sulfide group of an amino acid residue of the protein behaves as a nucleophile.

Glycinin is the most widespread storage protein in soybean, as well as one of the major allergens. The glycinin belongs to the globulins family, and as the Beta-conglycinin, each subunit is composed of an acidic and a basic chain. Pedersen et al.³⁵ reported IgE-binding to all of the native acidic subunits of glycinin but little or none to the basic subunits.

2S albumin is also a storage protein, but it belongs to the 2S seed storage albumins family and is formed by two chains linked by a disulfide bridge. As LTP, it is classified in the prolamin superfamily having a similar structure of alpha helices. The allergenicity of this protein also depends on this similarity.

Lectin is a protein found in the seeds of plants, mainly of leguminous plants. It is formed by four subunits, 30 kDa each,

and it is able to bind sugars. It is involved in carbohydrate recognition by a virus membrane and plays an important role in the immune system of plants. Barnett et al.³⁶ have investigated the possible role of lectin binding of IgE in RAST of legume and wheat extracts. They reported that there was significant immune binding of some of the lectins by specific IgE, and they concluded that these lectins may be important in expression of IgE-mediated allergic responses.

Trypsin inhibitor A belongs to the protease inhibitor I3 (leguminous Kunitz-type inhibitor) family, and it has endopeptidase and protease inhibitor activity. This protein is part of the defense system of the plant. Hypersensitivity to trypsin inhibitor is also documented by inhalation and was reported for the first time on soybean mill workers. Quirce et al.³⁷ have shown a relationship between a skin and a bronchial response to purified soybean KTI and work-related asthma in two bakers who were occupationally exposed to soybean flour.

Despite their well-documented allergenicity, soybean derivatives continue to be increasingly used in a variety of food products due to their well-documented health benefits. In this context, the knowledge of the protein profile of the soybean milk becomes important for a conscious consumption. We believe that proteomics should be considered a powerful tool in the functional characterization of the proteome as well as in the assessment of food safety. Indeed proteomic analysis could help to define specific allergenic proteins present in complex matrices, such as soybean seeds or soybean milk.

■ ASSOCIATED CONTENT

● Supporting Information

Proteins identified in soybean seeds, soybean milk, and soybean milk by precipitation; GO terms pie charts: GO annotation in biological process, cellular component and molecular function of the 243 nonredundant identified proteins and relative distribution of the seeds proteins, soybean proteins and precipitated soybean milk proteins; result of BLAST search in Uniprot. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: chiara.cavaliere@uniroma1.it. Fax: +39 06 490631. Phone: +39 06 49913834.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Consorzio Agrario dell'Emilia (Società Cooperativa, San Giorgio di Piano, BO, Italy), Dr. (Dott.) Stefano Monari and Dr. (Dott.ssa) Silvia Forapani for the kind donation of soybean seeds and soybean milk employed in this study.

■ ABBREVIATIONS USED

2D-PAGE, two-dimensional polyacrylamide gel electrophoresis; AGC, automatic gain control; DTT, dithiothreitol; fwhm, full width at half-maximum; GO, gene ontology; IAA, iodoacetamide; LTP, lipid transfer protein; PBs, protein bodies; RT, room temperature; SDS, sodium dodecyl sulfate; SPE, solid phase extraction; TCA, trichloroacetic acid; TFA, trifluoroacetic acid

REFERENCES

- (1) Paucar-Menacho, L. M.; Amaya-Farfán, J.; Berhow, M. A.; Mandarino, J. M. G.; Mejia, E. G.; Chang, Y. K. A high-protein soybean cultivar contains lower isoflavones and saponins but higher minerals and bioactive peptides than a low-protein cultivar. *Food Chem.* **2010**, *120*, 15–21.
- (2) Cavaliere, C.; Cucci, F.; Foglia, P.; Guarino, C.; Samperi, R.; Laganà, A. Flavonoid Profile in Soybeans by High Performance Liquid Chromatography–Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2177–2187, DOI: 10.1002/rcm.3049.
- (3) Derbyshire, E.; Wright, D. J.; Boulter, D. Legumin and vicilin storage protein of legume seeds. *Phytochem.* **1976**, *15*, 3–24.
- (4) Spector, D.; Anthony, M.; Alexander, D.; Arab, L. Soy consumption and colorectal cancer. *Nutr. Cancer.* **2003**, *47*, 1–12.
- (5) Badger, T. M.; Ronis, M. J. J.; Simmen, R. C. M.; Simmen, F. A. Soy protein isolate and protection against cancer. *J. Am. Coll. Nutr.* **2005**, *24*, 146S–149S.
- (6) IUIS Allergen Nomenclature Sub-Committee. <http://www.allergen.org/search.php?allergensource=Glycine+max> (accessed July 2014).
- (7) Kuppannan, K.; Julka, S.; Karnoup, A.; Dielman, D.; Schafer, B. 2DLC-UV/MS Assay for the Simultaneous Quantification of Intact Soybean Allergens Gly m 4 and Hydrophobic Protein from Soybean (HPS). *J. Agric. Food Chem.* **2014**, *62*, 4884–4892.
- (8) Holzhäuser, T.; Wackermann, O.; Ballmer-Weber, B. K.; Bindslev-Jensen, C.; Scibilia, J.; Perono-Garoffo, L.; Utsumi, S.; Poulsen, L. K.; Vieths, S. Soybean (Glycine max) allergy in Europe: Gly m (β -conglycinin) and Gly m 6 (glycinin) are potential diagnostic markers for severe allergic reactions to soy. *J. Allergy Clin. Immunol.* **2009**, *123*, 452–458.
- (9) Kuo, H. Y.; Chen, S. H.; Yeh, A. I. Preparation and Physicochemical Properties of Whole-Bean Soymilk. *J. Agric. Food Chem.* **2014**, *62*, 742–749.
- (10) Magni, C.; Ballabio, C.; Restani, P.; Sironi, E.; Scarafoni, A.; Poesi, C.; Duranti, M. Two-dimensional electrophoresis and Western-blotting analyses with anti Ara h 3 basic subunit IgG evidence the cross-reacting polypeptides of *Arachis hypogaea*, *Glycine max*, and *Lupinus albus* seed proteomes. *J. Agric. Food Chem.* **2005**, *53*, 2275–2281.
- (11) 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.
- (12) Hajdúch, M.; Ganapathy, A.; Stein, J. W.; Thelen, J. J. 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 Physiol.* **2005**, *137*, 1397–1419.
- (13) Mooney, B. P.; Thelen, J. J. High-throughput peptide mass fingerprinting of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. *Phytochem.* **2004**, *65*, 1733–1744.
- (14) Natarajan, S.; Xu, C.; Caperna, T. J.; Garret, W. M. Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. *Anal. Biochem.* **2005**, *342*, 214–220.
- (15) Gomes, L. S.; Senna, R.; Sandim, V.; Silva-Neto, M. A. C.; Perales, J. E. A.; Zingali, R. B.; Soares, M. R.; Fialho, E. Four Conventional Soybean [*Glycine max* (L.) Merrill] Seeds Exhibit Different Protein Profiles As Revealed by Proteomic Analysis. *J. Agric. Food Chem.* **2014**, *62*, 1283–1293.
- (16) Natarajan, S.; Luthria, D.; Bae, H.; Lakshman, D.; Mitra, A. Transgenic soybeans and soybean protein analysis: an overview. *J. Agric. Food Chem.* **2013**, *61*, 11736–11743.
- (17) 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*, 248–253.
- (18) Searle, B. C. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **2010**, *10*, 1265–1269.
- (19) Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, *74*, 5383–5392.
- (20) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646–4658.
- (21) Berg, M.; Parbel, A.; Pettersen, H.; Fenyő, D.; Björkstén, L. Reproducibility of LC-MS-based protein identification. *J. Exp. Bot.* **2006**, *57*, 1509–1514.
- (22) Murad, A. M.; Rech, E. L. NanoUPLC-MS^E proteomic data assessment of soybean seeds using the Uniprot database. *BMC Biotechnol.* **2012**, *12*, 82.
- (23) Miernyk, J. A.; Johnston, M. L. Proteomic analysis of the testa from developing soybean seeds. *J. Proteomics* **2013**, *89*, 265–72.
- (24) Roberts, R. C.; Briggs, D. R. Isolation and characterization of the 7S component of soybean globulins. *Cereal Chem.* **1965**, *42*, 71–85.
- (25) Hill, J. E.; Breidenbach, R. W. Proteins of soybean seeds. II. Accumulation of the major protein components during seed development and maturation. *Plant Physiol.* **1974**, *53*, 747–751.
- (26) Thanh, V. H.; Shibasaki, K. Heterogeneity of beta-conglycinin. *Biochim. Biophys. Acta* **1976**, *181*, 404–409.
- (27) Liu, H.; Sadygov, R. G.; Yates, J. R., III A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76*, 4193–4201.
- (28) Gonzalez, R.; Varela, J.; Carreira, J.; Polo, F. Soybean hydrophobic protein and soybean hull allergy. *Lancet* **1995**, *346*, 48–49.
- (29) Gonzalez, R.; Polo, F.; Zapatero, L.; Caravaca, F.; Carreira, J. Purification and characterization of major inhalant allergens from soybean hulls. *Clin. Exp. Allergy* **1992**, *22*, 748–755.
- (30) Quirce, S.; Polo, F.; Figueredo, E.; Gonzalez, R.; Sastre, J. Occupational asthma caused by soybean flour in bakers-difference with soybean-induced epidemic asthma. *Clin. Exp. Allergy* **2000**, *30*, 839–846.
- (31) Rihs, H. P.; Chen, Z.; Rueff, F.; Petersen, A.; Rozynek, P.; Heimann, H.; Baur, X. IgE binding of the recombinant allergen soybean profilin (rGly m 3) is mediated by conformational epitopes. *J. Allergy Clin. Immunol.* **1999**, *104*, 1293–1301.
- (32) Mittag, D.; Vieths, S.; Vogel, L.; Becker, W. M.; Rihs, H. P.; Helbling, A.; Wuthrich, B.; Ballmer-Weber, B. K. Soybean allergy in patients allergic to birch pollen: clinical investigation and molecular characterization of allergens. *J. Allergy Clin. Immunol.* **2004**, *113*, 148–154.
- (33) Ogawa, T.; Bando, N.; Tsuji, H.; Nishikawa, K.; Kitamura, K. α -Subunit of β -conglycinin, an allergenic protein recognised by IgE antibodies of soybean-sensitive patients with atopic dermatitis. *Biosci. Biotechnol. Biochem.* **1995**, *59*, 831–833.
- (34) Yaklich, R. W.; Helm, R. M.; Cockrell, G.; Herman, E. M. Analysis of the distribution of the major soybean seed allergens in a core collection of Glycine max accessions. *Crop Sci.* **1999**, *39*, 1444–1447.
- (35) Pedersen, H. S.; Djurtoft, R. Antigenic and allergenic properties of acidic and basic peptide chains from glycinin. *Food Agric. Immunol.* **1989**, *1*, 101–109.
- (36) Barnett, D.; Howden, M. E. H. Lectins and the radio-allergosorbent test. *J. Allergy Clin. Immunol.* **1987**, *80*, 558–561.
- (37) Quirce, S.; Fernandez-Nieto, M.; Polo, F.; Sastre, J. Soybean trypsin inhibitor is an occupational inhalant allergen. *J. Allergy Clin. Immunol.* **2002**, *109*, 178.