

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,300

Open access books available

130,000

International authors and editors

155M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Proteomics and Its Use in Obtaining Superior Soybean Genotypes

Cristiane Fortes Gris and Alexana Baldoni

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/51353>

1. Introduction

Soybean (*Glycine max* L. Merrill) is one of the most important and most cultivated crops in the world, with significant quantities of proteins being found in their yield composition, around 40% of their yield dry matter. This expressive quantity of proteins, and also a considerable percentage of oil, around 21% of their dry matter, has turned this grain into a product of great importance for the industrial sector, whether it be for food, cosmetics or, more recently, biofuels. Thus, soybean breeding programs directed toward these areas become ever more important, together with agronomic characteristics that allow greater productivity in sustainability with the environment in which they are produced.

The achievement of soybean genome sequencing [1], facilitated by identification of the genetic base, lead to advances in obtaining improved cultivars through knowledge of the complete sequence of expressed genes. Nevertheless, this information is not sufficient to identify which proteins are really being expressed in the cell at a given moment and under a certain condition since, through the phenomenon of splicing, different proteins may be produced by alteration of the command of a single gene. Thus, the complementary DNA (cDNA) and the messenger RNA (mRNA) have come to be the main focus of study for obtaining information regarding genetic expression or transcriptome. Nevertheless, due to post-translational regulation mechanisms, the quantity of expressed protein is not necessarily proportional to the quantity of its corresponding mRNA, which often raises questions regarding the role of this gene in cellular metabolism.

The reason for this is that control of gene expression occurs from mRNA transcription up to post-translational modifications like glycosylation and phosphorylation, among other processes, which alter protein activity (Figure 1).

In recent years, for the purpose of complementing the information obtained by means of genome sequencing and transcriptome, proteomics, one of the dimensions of the post-genome era [2], arises with a set of highly powerful techniques for separation and identification of proteins in biological samples, allowing better understanding of the networks of cellular operation and regulation upon representing the link between the genotype and the phenotype of an organism.

For the aforementioned reasons, proteomic analysis is now one of the most efficient means for functional study of the genes and genomes of complex organisms [3]. This has generated new data, as well as validated, complemented and even corrected information obtained through other approaches, thus contributing to better understanding of plant biology.

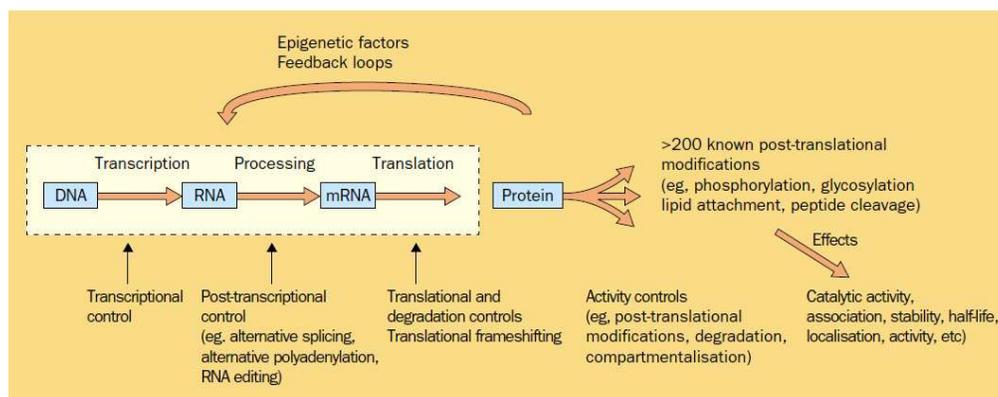


Figure 1. Pathways in which gene and protein expression may be regulated or modified in transcription or in post-translation [13].

Its study involves the entire set of proteins expressed by the genome of a cell, or only those that are expressed differentially under specific conditions. Also it is directed to the set of protein isoforms and post-translational modifications, to the interactions among them, as well as to the structural description of molecules and their complexes.

Bidimensional electrophoresis and mass spectrometry are the core technologies of proteomics, although new methodologies are being applied to plants for specific studies [4,5,6]. Among the most recent proteomic techniques are Difference Gel Electrophoresis (DIGE) and Multi-dimensional Protein Identification Technology (MudPIT), used in separation of proteins from a complex mixture. Other methods involved are Stable Isotopic Labeling using Amino Acids in Cell Culture (SILAC), Isotope Coded Affinity Tag (ICAT) and Isobaric Tag for Relative and Absolute Quantitation (iTRAQ) are based on labeling with isotopes for quantification of molecules by mass spectrometry.

In spite of the recent nature of research in this area, diverse studies with soybeans using proteomic tools are being performed throughout the world, showing this to be a promising area for selection of genotypes for genetic breeding programs [7,8]. Moreover, the study of plant responses to infections from pathogens has supplied significant data for understanding the signaling process that triggers the defense response in plants [9]. Additionally, there are

studies characterizing the proteome of plants in response to different stress conditions arising from both abiotic factors [10] and biotic factors [11]. These comparative studies of contrasting genotypes for a determined type of stress allow identification of the proteins that respond to stress by means of changes in their levels of expression. Identifying these molecules and their respective functions, the work of breeding is directed and should have continuity only with those molecules that perform roles related to the characteristic of stress tolerance. For that reason, it is essential to cross the proteomic data with information also obtained by genomics, transcriptomics and metabolomics, so as to verify the correlation of the candidate proteins with the desired characteristic.

In relation to products derived from genetically modified foods, proteomic techniques have been applied to allow a broad approach and the analysis of many variables simultaneously in a single sample. There are also other studies relating the proteome expressed during development of the plants, as well as research in which soybeans have been the target of investigations regarding nutritional, toxicological and allergenic aspects, above all on genetically modified varieties [12]. This makes for increased use of this technique in biosecurity studies. In this context, the objective of this chapter is to present the main technologies used in proteomic studies in diverse areas of activity, as well as the main scientific results obtained in the search for superior soybean genotypes.

2. Technologies used in proteomic studies.

Execution of a proteomic study involves the integration of many technologies which permeate the fields of molecular biology, biochemistry, physiology, statistics and bioinformatics, among other areas. The key steps in this type of study are separation of complex mixtures of proteins and their identification.

Separation is performed through the use of electrophoresis a term created by Michaelis in 1909. The first electrophoresis of proteins (Figure 2) was performed in 1937. Alfenas (1998) [14] explains that electrophoresis aims at separation of molecules in terms of their electrical charges, their molecular weights and their conformations, in porous supports and appropriate buffers, under the influence of a continuous electrical field. Molecules with a preponderance of negative charges migrate in the electrical field to the positive pole (anode), and molecules with excess of positive charges migrate to the negative pole (cathode). The preponderant charge of a proteic molecule is in accordance with its amino acids.

Many of the technologies currently used in proteomics were developed much before the beginning of proteomics, as is the case of electrophoresis. Nevertheless, it was the advance in protein sequencing technology by means of mass spectrometry that allowed its emergence and development [15].

The study of proteomics may be performed by means of techniques like two-dimensional electrophoresis in polyacrylamide gel (2D PAGE) followed by mass spectrometry (MS) (Figure 3), or furthermore, more recently, by the association of ionization and chromatographic

methods, among others, which increase detection sensitivity even more. Nevertheless, the point of departure has still been the exposure of a large number of proteins from a cell line or organism in two-dimensional polyacrylamide gels [16,17,18].

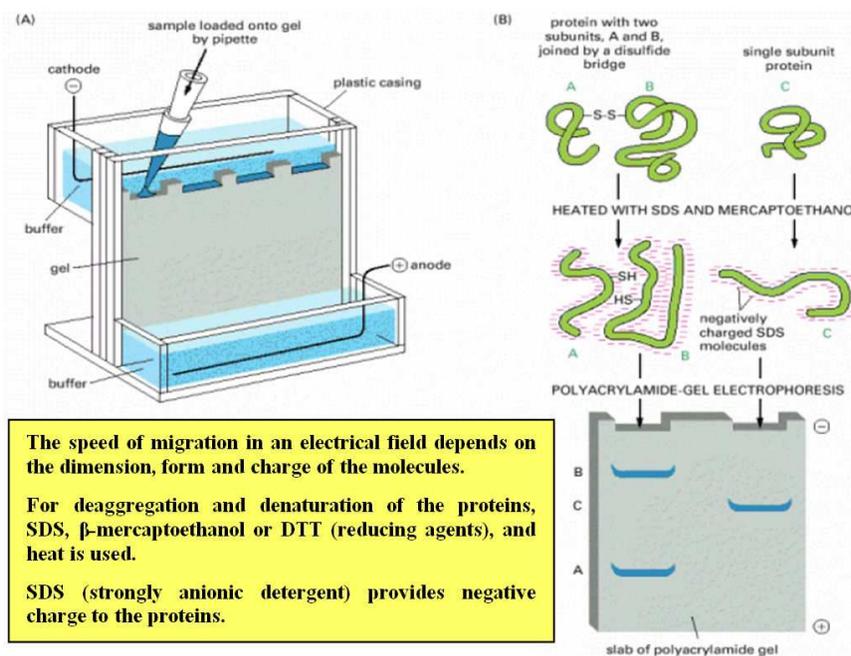


Figure 2. Polyacrylamide-gel electrophoresis (SDS-PAGE) used in proteome analysis [19].

2.1. Two-dimensional polyacrylamide gel electrophoresis (2D PAGE).

Two-dimensional polyacrylamide gel electrophoresis constitutes an analytical method capable of separating hundreds of proteins in a single analytical run. In this case, the gel, with the sample already applied, is submitted to an electrical field for two-dimensional separation. In the first dimension, separation occurs through isoelectric focalization, in which physical separation of the proteins occurs in terms of their respective isoelectric points on a strip of polyacrylamide with continuous gradation and known pH (IPG - immobilized pH gradient) submitted to increasing voltage. In the second dimension, the proteins under focus are submitted to polyacrylamide gel electrophoresis in the presence of SDS (SDS-PAGE) for separation according to their specific molecular masses (Figure 4). Thus, this is a technique that separates the proteins through different charges and masses.

The result of two-dimensional electrophoresis is a profile of spot distribution formed by single proteins or simple mixtures of proteins [21]. Each spot visualized in the gel may be considered as an orthogonal coordinate of a protein that migrated specifically in accordance with its isoelectric point (x axis) and its molecular mass (y axis), as shown in Figure 4.

The next step consists of staining the gel with silver, Coomassie blue, fluorescence, radioactive labeling or specific markers for phosphoproteins and glycoproteins, among others. This allows visualization of the protein expression pattern and photodocumentation of the gel

(Figure 5). After that, sectioning and digestion of selected spots of the gel are carried out and, finally, proteins of interest are identified by mass spectrometry integrated with a bioinformatics tool.

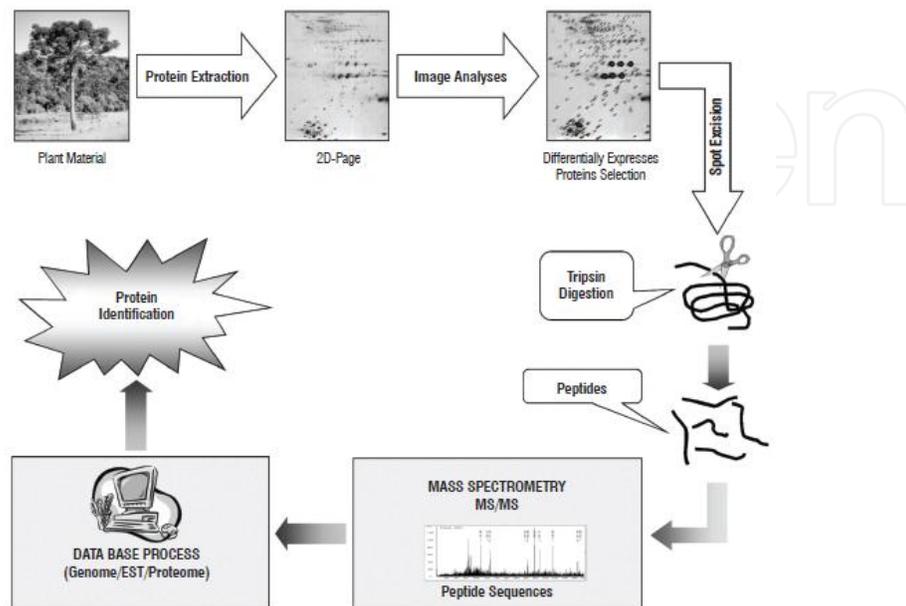


Figure 3. Stages of plant proteomics, using interface two-dimensional electrophoresis (2D-PAGE) and mass spectrometry [20].

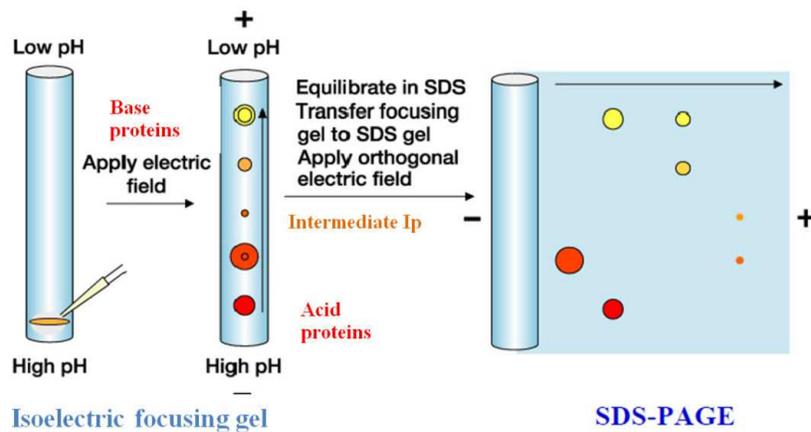


Figure 4. Two-dimensional electrophoresis 2D-PAGE used in analysis of proteomes [19].

Two-dimensional electrophoresis gels reflect the protein expression pattern of the biological sample analyzed and allow detection of variation of even a single amino acid between two isoforms or covalent modifications in the same protein thanks to change in the position of the spot.

It is important to highlight that each sample, depending on its nature, requires a specific type of processing for extraction and focalization. Therefore, it is expected that the user checks beforehand in related publications as to the protocols and methodologies that best suit the experimental needs.

Some limitations are associated with two-dimensional electrophoresis, such as low reproducibility and little power of automation. Nevertheless, reproducibility may be increased by defining optimal conditions for the electrophoresis, while automation of the process is only possible in relation to analysis of gels. Gel analysis software determines the spots and identifies those expressed differentially and their volumes, inferring a relative quantification of expression of that protein in comparison to the same spot of another gel [22]. Thus, by a process of subtraction, the differences among the different samples are revealed, as, for example, the presence, absence or intensity of the proteins. Thus, the proteins of interest may then be identified based on knowledge of the isoelectric point and of apparent molecular weight, determined by the two-dimensional gels [23].

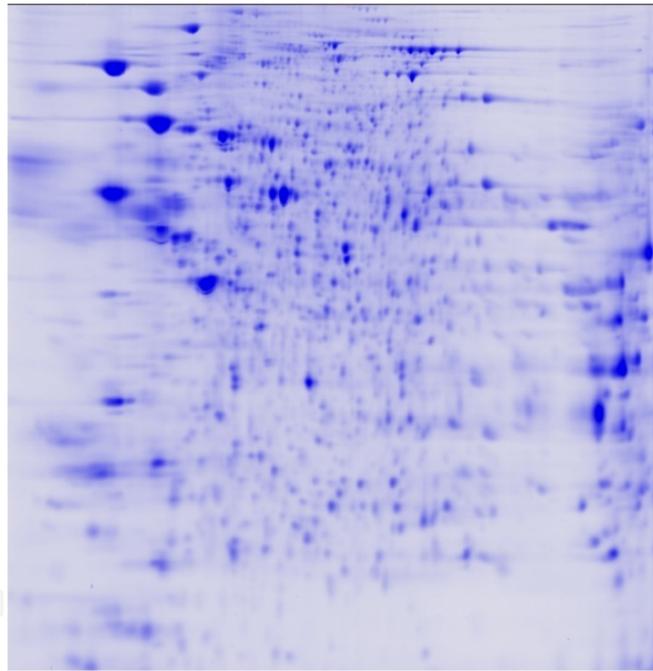


Figure 5. Proteins extracted and separated by two-dimensional (2D) gel electrophoresis and stained with Coomassie blue [24].

2.2. Differential in gel electrophoresis (DIGE).

An efficient procedure in the attempt to eliminate variation from gel to gel is use of the technique of differential in gel electrophoresis or DIGE (Figure 6), which allows analysis of up to three proteomes in a single gel. These results in one internal pattern common to all the gels and two different samples labeled with distinct fluorophores (CyDye) [25]. That way, only the proteins labeled with their own fluorophore are visualized. In addition, this technique

uses labeling of proteins with a broad dynamic range of detection and has sensitivity greater than staining of the gels by silver methods, allowing proteomic studies of a quantitative nature to be performed with greater precision, accuracy and sensitivity [26].

2.3. Liquid chromatography

Another form used for separation of proteins is by means of liquid chromatography. The sample that is, for example, a mixture of peptides generated by proteolytic digestion from a protein extract passes through a first separation, by means of liquid chromatography, where the enriched peptide fractions are collected and applied in the spectrometer. As complete automation is the main target of the methods for large scale analyses, methods of separation were developed free of gel by reverse phase liquid chromatography connected with tandem mass spectrometry (LC/MS/MS). In Figure 7 the operational and equipment sequence involved in a typical analysis via LC/MS/MS is shown.

Greater automation is possible with multidimensional liquid chromatography, which uses different characteristics of the proteins in columns of distinct properties or in a single two-phase column [29]. The fraction eluted in the first column is directly introduced in the second column, which may be directly connected to the mass spectrometer. This technique, called MudPIT, is inserted in the context of the shotgun proteomic, in which greater resolution of the proteomes is possible, facilitating identification of the less abundant proteins frequently lost when gels are used [30].

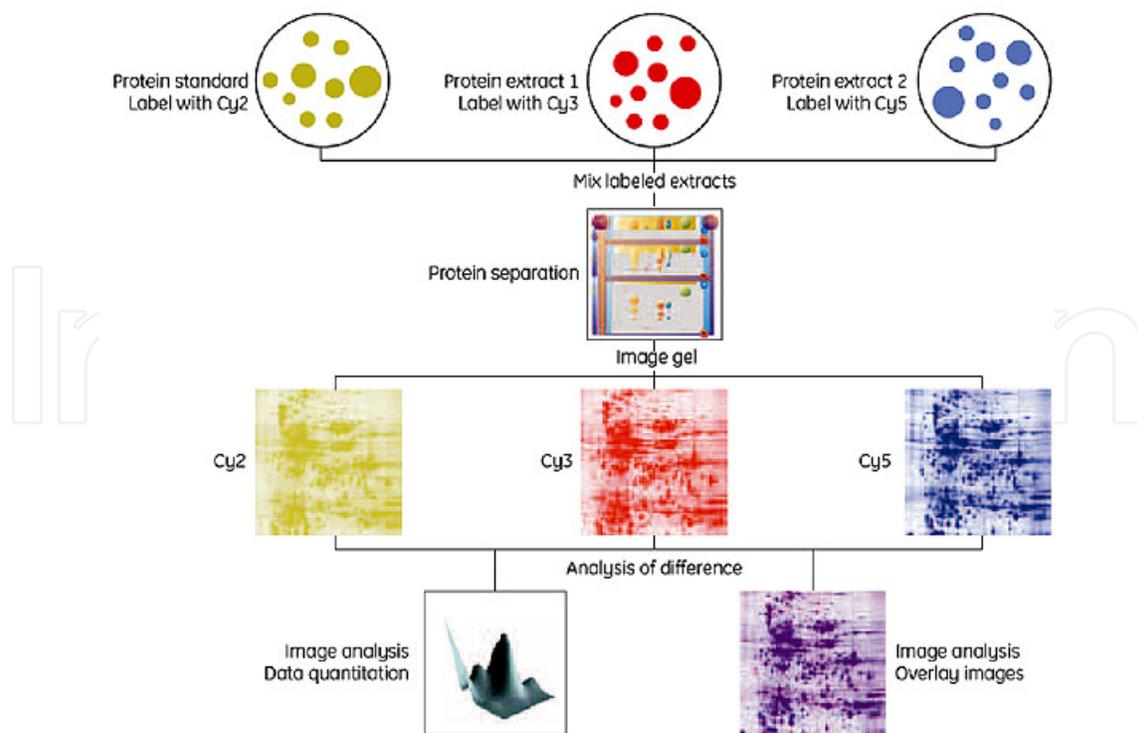


Figure 6. Differential in gel electrophoresis technique or DIGE [27].

Protein Identification with Chromatographic Separation (LC/MS/MS)

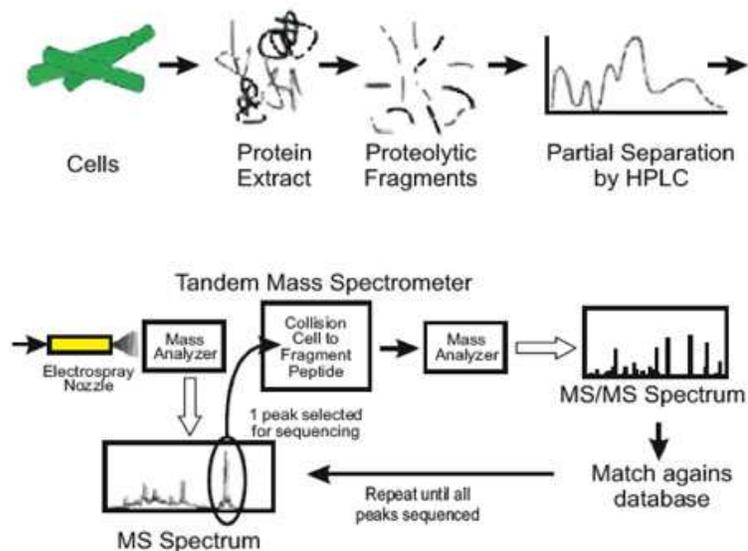


Figure 7. Protein identification with chromatographic separation (LC/MS/MS) [28].

2.4. Protein identification methods.

After separation of proteins, the next stage consists of their characterization and identification using mass spectrometry, which is a technique where the ratio between the mass and the charge (m/z) of ionized molecules in the gas phase is measured. In general, a mass spectrometer consists of an ionization source, a mass analyzer, a detector and a data acquisition system.

The great variety of spectrometers found on the market is the result of different combinations of types of sources of ionization and mass analyzers, which provide certain levels of sensitivity and accuracy in the results. At the ionization source, the molecules are ionized and transferred to the gas phase. In the mass analyzer, the ions formed are separated in accordance with their m/z ratios and later detected, usually by electron multiplier [31].

With the development of ever more specialized equipment for proteins, mass spectrometry has become a revolutionary tool in modern protein chemistry. This technology has allowed identification of proteins by a methodology called peptide mass fingerprinting. Rocha et al. (2003) [3], state that this methodology is based on protein digestion to be identified by a proteolytic enzyme, for example trypsin, producing fragments called peptides. The masses of these peptides obtained form a kind of fingerprinting of the protein, which are then determined with great acuity (0.1 to 0.5 Da) by mass spectrometry.

Special software allows comparing the peptide mass fingerprinting of the protein one wishes to identify with those theoretically generated for all the protein sequences present in the databases. If the protein sequence problem is in the database, it will immediately be identified [32].

2.5. Relative protein quantification

Large scale protein quantification methods make an estimate of relative expression possible by means of labeling with radioactive isotopes, fluorescents and light/heavy, allowing the same protein to be quantified in a relative way among differently labeled samples. Some of the most used radioactive isotopes are the iCAT (Isotopic coded affinity tag), iTRAQ (isobaric tags) and H_2O^{18} .

The iCAT consists of addition of a label that has affinity for cysteine residues and which has a bonded molecule of eight atoms of hydrogen or eight atoms of deuterium. One sample is labeled with the tag containing hydrogen and the other sample with the tag containing deuterium. After digestion of the proteins, the resulting peptides are identified by mass spectrometry. Equal peptides labeled in the two samples are identified by overlap of the peaks that show distinct m/z due to the type of bonded isotope, with the ratio between the area of the two peaks being a relative measure of the expression of that protein. According to Yi & Goodlett (2003) [33], the main problems associated with this technique are the need for the presence of cysteine residues, the high cost of the reagents and the greater time necessary for sequencing.

In the iTRAQ technique, labeling of proteins with tags and identification by mass spectrometry is also used. The tags bond to all the free amino groups at the N terminal of all the peptides and on the internal side chains with lysine residues and vary according to the reporter group they carry, and they may have 114, 115, 116 or 117Da, thus allowing for the quantification of proteins in up to four types of samples at the same time. The relative quantification is carried out in the same way as in the iCAT, but high cost has restricted its use [34].

The aforementioned techniques require the consumption of specific and expensive reagents. Nevertheless, the same goal may be achieved with a simpler labeling method in which the proteins are labeled with one or two atoms of O_2 . These are incorporated in the carboxyl terminal by simply supplying a solution with H_2O for one sample and a solution with H_2O^{18} for the other sample. Thus, the relative abundance of the peptides that will differ by 2Da is estimated [35].

Another quantification technique is Stable isotope labeling by amino acids in cell culture, (SILAC) which, together with mass spectrometry and bioinformatics resources, has proven to be quite adequate in proteomic studies. It is a technique that detects differences in the abundance of proteins among cell cultures by means of isotopic labeling of proteins. Labeling with stable isotopes is obtained by supplying isotopically enriched amino acids to a cell culture and natural amino acids to the culture to be compared (Figure 8).

2.6. Analysis of post-translational modifications (PTM's).

Another area of great interest in plant proteomics is in regard to characterization of post-translational modifications or PTM's, essential for proteins to play their roles in the varied cell events, producing different proteins from the same gene.

These modifications occur at specific sites in the proteins [37] changing their physical, chemical and biological properties [38]. They may occur by means of cleavages or by the addition

of a chemical group to one or more amino acids [39]. The main goals of PTM studies in proteomics are identifying the proteins that have them, mapping the sites where these modifications occur, quantifying their occurrence at the different sites and characterizing cooperative PTM's [40].

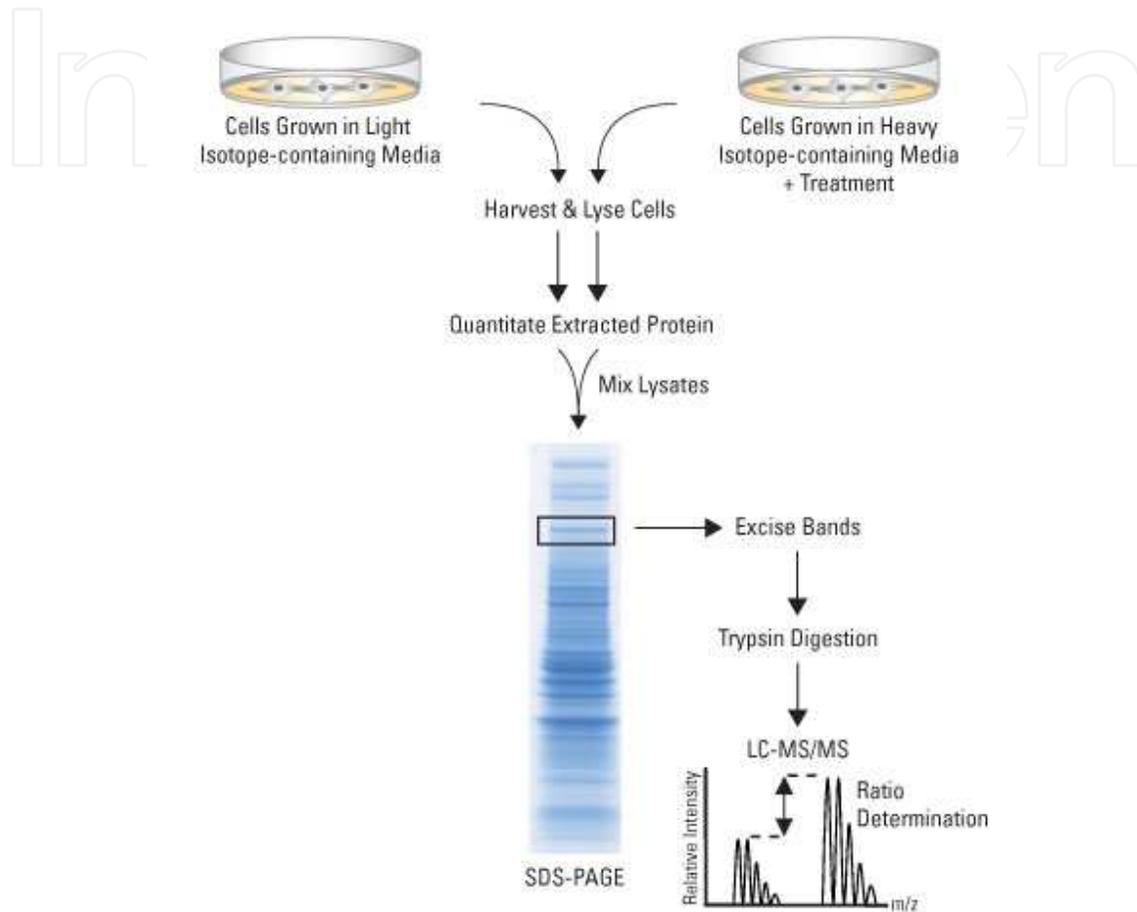


Figure 8. General outline of the SILAC technique [36].

The fact that covalent modifications result in changes in the protein molecular masses makes it possible for these modifications and the amino acids that carry them to be identified by mass spectrometry, allowing more than 300 different types of PTM's to be identified until now with the aid of this technique. Nevertheless, according to Mann and Jensen (2003) [41], mass spectrometry has reduced power of resolution of PTM's because they occur at low stoichiometric levels. This problem may be resolved by adopting fractionating methods prior to sequencing that allow enrichment of the sample for the proteins that have a certain type of PTM. Large scale modified protein enrichment systems are generally carried out by means of affinity chromatography.

One example is the IMAC system – a column of immobilization through affinity to a metal for isolation of phosphorylated proteins in which metal ions of Fe(III) are joined to the ma-

trix to promote the isolation of proteins that have phosphorylate residues since the Fe(III) ion is capable of interacting in a reversible manner with the phosphate group of the modified peptide keeping it attached to the column [41].

Contrary to that which occurs with the reversible yet permanent PTM's, like glycosylation, low stoichiometry does not occur, but the addition of carbohydrates hinders the proteolytic digestion necessary for identification by mass spectrometry [21]. In addition, when the modified peptide is fragmented for sequencing, it loses sugar residues, impeding the identification of the modified amino acids. To resolve this problem, digestion of the proteins is performed so as to remove the sugar residues and produce a modification in the modified site that makes it identifiable [42].

Electrophoresis gels may also be used in enrichment of samples for PTM's as performed for detection of phosphorylations and glycosylations with commercially available kits. The modified proteins, specifically labeled in the gel, are visualized and excised for identification by mass spectrometry. One important aspect of the use of gels for identification of PTM is the possibility of visualizing the spots differentially expressed among samples that have the PTM.

3. Research dealing with proteomics in soybeans.

3.1. Food safety

In the case of food, proteins are especially important for evaluation of food safety because they may place consumer health at risk. That is because proteins may be involved in synthesis of toxins and antinutrients, as well as being a toxin, an antinutrient or even an allergenic [43].

Soybeans are an important source of food throughout the world, being consumed in daily meals of all types. It has also been widely used as a food substitute by people that have intolerance to lactose or other milk proteins [44]. Nevertheless, in this species are also found proteins considered allergenic. Thus, knowledge regarding the proteins with toxic/antinutritional potential present in this grain becomes fundamental for development of biotechnological strategies that would have the target of elimination or inactivation in the genome of these species of genes that codify for these proteins.

Therefore, application of proteomic analysis in this type of study has been widely discussed. In relation to products derived from genetically modified (GM) foods, proteomic techniques have been applied because they allow a wide-ranging approach and analysis of many variables simultaneously in the same sample [45]. Ocana et al. (2007) [46], studying GM proteins present in soybean and maize samples using proteomic analysis, identified the protein CP4 EPSPS, which confers tolerance to glyphosate herbicide. These samples were submitted to specific separation techniques followed by two-dimensional electrophoresis and mass spectrometry for detection and characterization of the proteome.

Related to allergies, various allergens belonging to the superfamily of cupins and prolamins have been identified in soybeans [47]. Research has suggested that a heterogeneous group of soybean proteins bond to the IgE antibody and are potential allergens as, for example, Gly

m Bd 30k, β -conglycinin, Gly m Bd 28k, glycinin, Kunitz type protease inhibitor, some proteins present in the hull (Gly m 1.0101, Gly m 1.0102 e Gly m 2), profilin (Gly m 3), SAM 22 (Gly m 4), and other allergens like lectin and lipoxygenase [47,48]. According to Wilson et al. (2005) [49], in spite of the allergens identified in soybeans, the challenge of food researchers is developing a process for eradicating the immunodominant allergens, maintaining the functionality, nutritional value and effectiveness in the subsequent products derived from soybeans. For that reason, research has been developed using genetic engineering for silencing the soybean gene responsible for synthesis of the protein Gly m Bd 30K, one of the main soybean proteins that develop allergic reactions with serums of sensitive patients [44].

3.2. Biotic and abiotic factors

In a similar manner, various studies have shown that the proteomic approach is highly useful for investigation of crop response to environmental stresses because it compares the way the proteome is affected by different physiological conditions.

Saline stress is one of the many types of abiotic stresses that affect plants and compromise their yield. Salinity is a common agricultural problem in arid and semiarid regions and creates large unproductive areas. There has been an ever greater search for cultivars adaptable to this condition. Sobhanian et al. (2010) [10], used proteomic techniques to evaluate the metabolism of proteins in leaves, hypocotyls and roots submitted to different NaCl concentrations (Figure 9), thus leading to saline stress.

Results in soybeans suggest that, in adaptation to saline conditions, proteins perform different roles in each organ, and the proteins most affected by saline stress are those related to photosynthesis. Therefore, there is less energy production, and, consequently, reduction in plant growth. The conclusion suggests that the gene Glyceraldehyde-3-phosphate dehydrogenase may be, in the future, one of the target genes to improve tolerance to saline stress in this species.

Another type of abiotic stress studied in soybeans in which a proteomic approach is used is flooding stress [50,51]. Growing this species in areas subject to flooding makes the root environment anoxic, affecting nodulation or root growth. That way, plants respond with greater or less efficiency, allowing the distinction between cultivars which are tolerant and intolerant to this stress.

Proteomic analyses of soybean seedlings in response to flooding were undertaken by Shi et al. (2008) [52] to identify the key proteins involved in this process. To identify the first proteins produced in response to flooding, the roots of the seedlings were used for extraction of the proteins. The two-dimensional gel results suggest that cytosolic ascorbate peroxidase 2 (cAPX 2) is involved in response to flooding stress in young soybean seedlings.

In the case of drought stress, up-regulation of reactive oxygen species (ROS) scavengers such as superoxide dismutase (SOD) was reported in soybean seedlings [53]. The proteome analysis of two-day-old soybean seedlings subjected to drought stress by withholding of water for two days revealed a variety of responsive proteins involved in metabolism, disease/defense and energy including protease inhibitors [53]. The major reason for loss of crop yields under drought stress is a decrease in carbon gain through photosynthesis. Proteome

analysis of soybean root under drought condition showed that two key enzymes involved in carbohydrate metabolism, UDP- glucose pyrophosphorylase and 2,3-bisphosphoglycerate independent phosphoglycerate mutase, were down-regulated upon exposure to drought [54]. The identification of proteins such as UDP-glucose pyrophosphorylase and 2,3- bisphosphoglycerate has provided new insights that may lead to a better understanding of the molecular basis of responses to drought stress in soybean

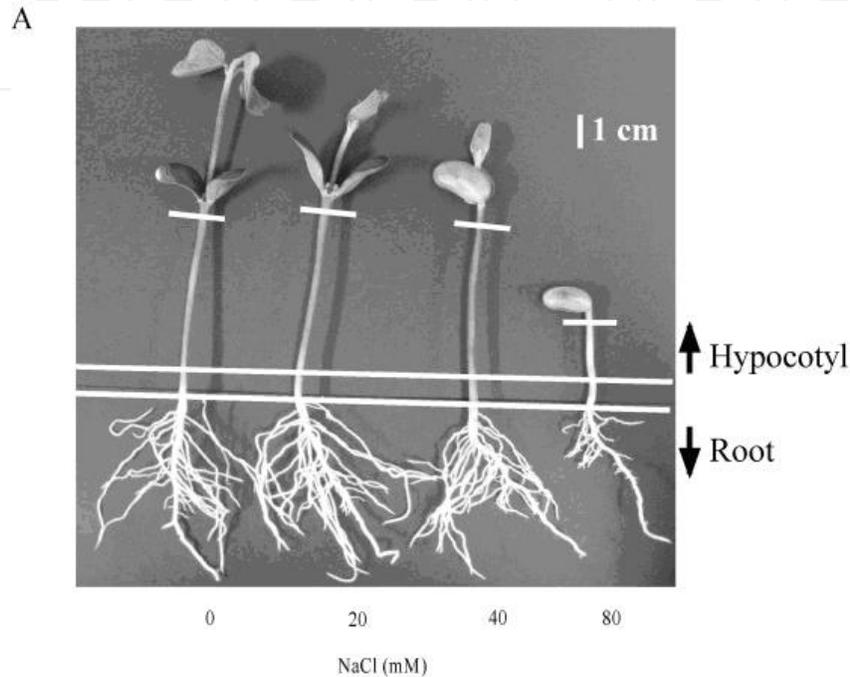


Figure 9. Soybean seedlings submitted to different concentrations of NaCl [10].

Stress by toxicity caused by the presence of high quantities of aluminum in the soil has also been investigated in soybeans from the perspective of proteomics [55,56]. Duressa et al. (2011) [56], studying cultivars tolerant and susceptible to high doses of aluminum, made proteomic analyses of roots, arriving at the conclusion that the greatest expression of enzymes involved with citrate synthesis would be a good strategy in the search for cultivars tolerant to this mineral (Figure 10).

Another focus of the study within the context of selection of superior soybean genotypes using the proteomic approach is exposure to ultraviolet radiation, which has gained importance with the prominent worldwide concern for global warming and the consequent degradation of the ozone layer. Xu et al. (2007) [57], studied the proteome of soybean leaves to investigate the protective role of flavonoids against the incidence of UV-B radiation. The authors suggest that high levels of flavonoid reduce the sensitivity of the plant to this radiation.

In relation to biotic stresses caused by pathogens like fungi, bacteria, nematodes and viruses, proteomic tools are also greatly used because they allow understanding of the plant-pathogen relationship [11,58,59,60] and also how the nodulation process occurs by means of symbiosis between the soybean roots and rhizobia [61]. In these cases, proteomic analysis

provides the information that will be used by genetic breeding in the search for cultivars resistant to various diseases.

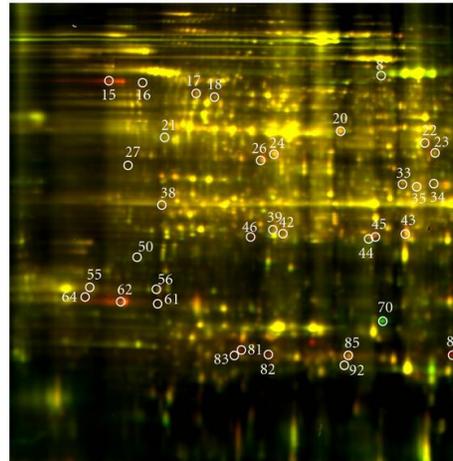


Figure 10. profile of aluminum regulated-proteins in PI 416937 72 h posttreatment [56].

Zhang et al. (2011) [58] evaluated the responses of cultivars tolerant and susceptible to the fungus *Phytophthora sojae* by means of two-dimensional electrophoresis. The authors observed 46 proteins being expressed (Figure 11), among which only 11% were related to plant defense.

In addition, proteomic studies that deal with seed development also play an essential role [62]. The data obtained may help to interpret the function of genes that determine protein concentration, considered as a key characteristic for genetic breeding of soybeans. Moreover, differential proteomic analyses designed to describe the changes that occur from maturation to senescence in organs and organelles have been reported. There is also already a soybean proteome database, providing information on the proteins involved in the soybean response to stress caused by drought, salinity and, principally, flooding [63].

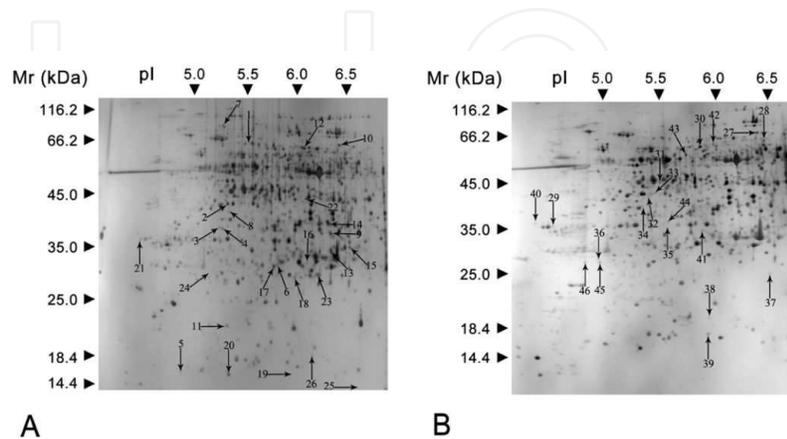


Figure 11. Identification of 26 and 20 protein spots from Yudou25 (A) and NG6255 (B), respectively. The numbers with arrows indicate the differentially expressed protein spots. Ip and Mr are shown on the gels [58].

4. Final considerations

In light of the above, proteomics in soybean studies contributes to diverse biotechnological applications, with its approach proving to be fundamental. Its use in the search for superior soybean materials has the purpose of comparing and contrasting genotypes for a determined type of stress and identifying the proteins that respond to the stress by means of changes in their levels of expression. The identification of these molecules and their respective functions will allow direction of breeding work, which should continue only with those that perform roles related to the characteristic of stress tolerance.

For that reason, it is essential to cross proteomic data with information also gathered from genomics, transcriptomics and metabolomics so as to check the correlation of the candidate proteins with the desired characteristic. The following stage aims to evaluate these proteins (genes) in regard to their segregation for the characteristic of interest or quantitative trait locus (QTL), that is, determine how much each one of them contributes to the characteristic of tolerance. Finally, the selected genes may be integrated in marker assisted selection (MAS) or in genetic transformation programs.

Author details

Cristiane Fortes Gris¹ and Alexana Baldoni²

1 Federal Institute of Southern Mines

2 Federal University of Lavras, Brazil

References

- [1] Schmutz, J., Cannon, S. B., Schlueter, J., J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., et al. (2010). Genome sequence of the paleopolyploid soybean. *Nature*, 463, 178-183, <http://www.nature.com/nature/journal/463n7278/full/nature08670.html>, (accessing 20 february 2012).
- [2] Pandey, A., & Mann, M. (2000). Proteomics to study genes and genome. *Nature* .<http://65.199.186.23/nature/journal/n6788/abs/405837a0.html> (accessing 20 february 2012) , 405(6788), 837-846.
- [3] Rocha, T. L., Costa, P. H. A., Magalhaes, J. C. C., Evaristo, R. G. S., Vasconcelos, E. A. R., Coutinho, M. V., Paes, N. S., Silva, M. C. M., & Rossi-de-Sa, M. F. (2005). Eletroforese bidimensional e análise de proteomas. *Comunicado Técnico. Embrapa Recursos Genéticos e Biotecnologia*, 136, 1-12.

- [4] Mann, M. (2006). Functional and quantitative proteomics using SILAC. *Nature Review Molecular Cell Biology*, 7(12), 952-958, http://patf.lf1.cuni.cz/proteomika/silac_2006.pdf, (accessed 07 march 2012).
- [5] Mcdonald, T., Sheng, S., Stanley, B., Cheng, D., Ko, Y., Cole, R. N., Pedersen, P., & Van Eyk, J. E. (2006). Expanding the subproteome of the inner mitochondria using protein preparation technologies. One- and two-dimensional liquid chromatography and two-dimensional gel electrophoresis. *Molecular & Cellular Proteomics*, 5(12), 2392-2411, <http://www.mcponline.org/content/5/12/2392.full>, (accessed 05 march 2012).
- [6] Maor, R., Jones, A., Nuhse, T. S., Studholme, D. J., Peck, S. C., & Shirasu, K. (2007). Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. *Molecular & Cellular Proteomics*, 6, 601-610, <http://www.mcponline.org/content/6/4/601.full>, (accessed 2 march 2012).
- [7] Natarajan, S., Xu, C., Caperna, T. J., & Garrett, W. M. (2005). Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. *Analytical Biochemistry*, 342(2), 214-220, <http://www.sciencedirect.com/science/article/pii/S0003269705003593>, (accessed 2 march 2012).
- [8] Krishnan, H. B., & Nelson, R. L. (2011). Proteomic analysis of high protein soybean (*Glycine max*) accessions demonstrates the contribution of novel glycinin subunits. *Journal of Agricultural and Food Chemistry*, 59, 2432-2439, <http://pubs.acs.org/doi/abs/10.1021/jf104330n>, (accessed 2 march 2012).
- [9] Qureshi, M. I., Qadir, S., & Zolla, L. (2007). Proteomics-based dissection of stress response pathways in plants. *Journal of plant physiology*, 01, 1-13, <http://www.sciencedirect.com/science/article/pii/S0176161707000995>, (accessed 3 march 2012).
- [10] Sobhanian, H., Razavizadeh, R., Nanjo, Y., Ehsanpour, A. A., Jazii, F. R., Motamed, N., & Komatsu, S. (2010). Proteome analysis of soybean leaves, hypocotyls and roots under salt stress. *Proteome Science*, 8, 19-33, <http://www.biomedcentral.com/1477-5956/8/19>, (accessed 26 february 2012).
- [11] Liu, D., Chen, L., & Duan, Y. (2011). Differential proteomic analysis of the resistant soybean infected by soybean cyst nematode. *Heterodera glycines race 3*. *Journal of Agricultural Science*, 3(4), 160-167, <http://www.ccsenet.org/journal/index.php/jas/article/view/9658>, (accessed 26 february 2012).
- [12] Castro, V. A. O. T., & Finardi, F. F. (2008). Análise Proteômica de amostras de soja GM e parental. I XXI Congresso Brasileiro de Ciência e Tecnologia de Alimentos e XV Seminário Latino Americano e do Caribe de Ciência e Tecnologia de Alimentos, Belo Horizonte, Brasil XXI SBCTA, 2008.
- [13] Banks, R. E., Dunn, M. J., Hochstrasser, D. F., Sanchez, J., Blackstock, W., Pappin, D. J., & Selby, P. J. (2000). Proteomics: new perspectives, new biomedical opportunities. *The Lancet*, 356, 1749-1756, <http://cmbi.bjmu.cn/cmbidata/proteome/reviews/02.pdf>, (accessed 26 february 2012).

- [14] Alfenas, A.C. (1998). Eletroforese de isoenzimas e proteínas afins: fundamentos e aplicações em plantas e microorganismos. *Viçosa, MG: UFV.*, 574.
- [15] Tyers, M., & Mann, M. (2003). From genomics to proteomics. *Nature*, 422(6928), 193-197, <http://www.ncbi.nlm.nih.gov/pubmed/12634792>, (accessed 2 march 2012).
- [16] Anderson, N. G., & Anderson, N. L. (1996). Twenty years of Two-dimensional electrophoresis: past, present and future. *Electrophoresis*, 17, 443-453, http://siba.unipv.it/fisica/articoli/E/Electrophoresis1996_17_443.pdf, (accessed 26 february 2012).
- [17] Wilkins, M. R., Sanchez, J. C., Williams, K. L., & Hochstrasser, D. F. (1996). Current challenges and future applications for protein maps and posttranslational vector maps in proteome projects. *Electrophoresis*, 17(5), 830-838, <http://onlinelibrary.wiley.com/doi/10.1002/elps.1150170504/abstract>, (accessed 12 mai 2012).
- [18] Wilkins, M. R., Willians, K. L., Appel, R. D., & Hochstrasser, D. (1997). *Proteome research: new frontiers in functional genomics.*, Germany, Springer-Verlag, 243.
- [19] Stracham, T., & Read, A. P. (2004). *Human Molecular Genetics 3.*, New York, Garland Science.
- [20] Balbuena, T. S., Dias, L. L. C., Martins, M. L. B., Chiquieri, T. B., Santa-Catarina, C., Floh, E. I. S., & Silveira, V. (2011). Challenges in proteome analyses of tropical plants. *Brazilian Journal of Plant Physiology*, 23(2), http://www.scielo.br/scielo.php?pid=S167704202011000200001&script=sci_arttext, (accessed 09 march 2012).
- [21] Pennington, S. R., & Dunn, M. J. (2001). *Proteomics: from protein sequence to function*, New York, Springer-Verlag e BIOS scientific Plubishers, 1v.
- [22] López, J. L., Marina, A., Vázquez, J., & Alvarez, G. (2002). A proteomic approach to the study of the marine mussels. *Mytilus edulis* and *M. galloprovincialis*. *Marine Biology*, 141, 217-223, <http://www.whoi.edu/science/B/people/mhahn/Lopez.pdf>, (accessed 13 april 2012).
- [23] James, P. (1997). Protein identification in the post-genome era: the rapid rise of proteomics. *Quarterly reviews of biophysics*, 30, 279-331.
- [24] Applied Biomics: 2D Gel Staining. (2012). http://www.appliedbiomics.com/proteomics_differentstaining.html, (accessed 10may).
- [25] Unlu, M., Morgan, M. E., & Minden, J. S. (1997). Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis*, 18(11), 2071-2077, <http://onlinelibrary.wiley.com/doi/10.1002/elps.1150181133/abstract>, (accessed 15 april 2012).
- [26] Tong, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., & Davison, M. (2001). Validation and development of fluorescent two-dimensional gel electrophoresis proteomics technology. *Proteomics*, 1, 377-396, <http://www.ncbi.nlm.nih.gov/pubmed/11680884>, (accessed 12 april 2012).

- [27] Healthcare, G.E. (2012). *Life Sciences 2D Electrophoresis Principles and Methods*, https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1335426794335/litdoc80642960_20120426103127.pdf, (accessed 10 april 2012).
- [28] Iowa State University. (2012). Q-Star Tandem Mass Spectrometry. <http://www.protein.iastate.edu/q-star.html>, (accessed 15 april).
- [29] Motoyama, A., & Yates, J. R. (2012). Multidimensional L.C. separations in shotgun proteomics. *Analytical Chemistry*, 80(19), 7187-7193, <http://pubs.acs.org/doi/full/10.1021/ac8013669>, (accessed 12 mai 2012).
- [30] Washburn, M. P., Wolters, D., & Yates, J. R. (2008). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, 19, 242-247, http://proteome.gs.washington.edu/classes/Genome490/papers/Washburn_et_al_Nat_Biotech_2001.pdf, (accessed 28 april 2012).
- [31] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246, 64-71, <http://www.sciencemag.org/content/246/4926/64.abstract>, (accessed 25 february).
- [32] De Sousa, M. V., Fontes, W., & Ricart, C. A. O. (2003). Análise de Proteomas: O despertar da era pós-genômica. *Revista on line-Biotecnologia Ciência e Desenvolvimento*, 7, 12-14, <http://www.biotecnologia.com.br/revista/bio07/analise.pdf>, (accessed 10 april 2012).
- [33] Yi, E. C., & Goodlett, D. R. Quantitative protein profile comparisons using the isotope-coded affinity tag method. *Current Protocols in Protein Science*, 23, 2.1-23.2.11, <http://onlinelibrary.wiley.com/doi/10.1002/0471140864.ps2302s34/abstract?systemMessage=Wiley+Online+Library+will+be+disrupted+on+9+June+from+10%3A00-12%3A00+BST+%2805%3A00-07%3A00+EDT%29+for+essential+maintenance>, (accessed 28 april 2012).
- [34] Schmidt, C., & Urlaub, H. (2009). Itraq-labeling of in-gel digested proteins for relative quantification. *Methods in Molecular Biology*, 564(4), 207-226, <http://www.springerlink.com/content/t864281256610gr4/#section=78222&page=1&locus=0>, (accessed 28 april 2012).
- [35] Ye, X., Luke, B., Andresson, T., & Blonder, J. (2009). O stable isotope labeling in MS-based proteomics. *Briefings in Functional Genomics & Proteomics*, 8(2), 136-144, <http://bfg.oxfordjournals.org/content/8/2/136.full>, (accessed 15 april 2012).
- [36] Thermo Scientific: Pierce Protein Biology Products. (2012). SILAC Protein Quantitation Kits. <http://www.piercenet.com/browse.cfm?fldID=01040418>, (accessed 10 may).
- [37] Blom, N. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, 4(6), 1633-1649, <http://onlinelibrary.wiley.com/doi/10.1002/pmic.200300771/abstract>, (accessed 15 april 2012).
- [38] Nesaty, V. J., & Suter, M. J. F. (2008). Analysis of environmental stress response on the proteome level. *Mass Spectrometry Reviews*, 27(6), 556-574, <http://onlineli->

- brary.wiley.com/doi/10.1002/mas.20177/abstract;jsessionid=C9FCCA3D6997D990B13A7CAE38CB4F3E.d03t02 , (accessed 05 mai 2012).
- [39] Mann, M., & Jensen, O. N. (2003). *Nature Biotechnology*, 21, 255-261, <http://genomics.wsu.edu/pages/teaching/MPS570%20PMB/ArticlesPDFs/MannandJensen.pdf>, (accessed 02 april 2012).
- [40] Seo, J., & Lee, K. J. (2004). Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *Journal of Biochemistry and Molecular Biology*, 37(1), 35-44, http://bmbreports.org/jbmb/jbmb_files/%5B37-1%5D0401271834_035-044.pdf, (accessed 12 april 2012).
- [41] Vener, A. V., Harms, A., Sussman, M. R., & Vierstra, R. D. (2001). Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of *Arabidopsis thaliana*. *Journal of Biological Chemistry*, 276(10), 6959-6966, http://www.ufv.br/dbv/pgfvg/BVE684/htms/pdfs_revisao/genomica/protome_articles/Mass%20Spectrometric%20Resolution%20of%20Reversible%20Protein%20Phosphorylation.pdf, (accessed 10 april 2012).
- [42] Cantin, G. T., & Yates, J. R. (2004). Strategies for shotgun identification of post-translational modifications by mass spectrometry. *Journal of Chromatography A*, 1053(1-2), 7-14, <http://www.sciencedirect.com/science/article/pii/S0021967304009719>, (accessed 25 march 2012).
- [43] Ruebelt, M. C., Leimgruber, N. K., Lipp, M., Reynolds, T. L., Nemeth, M. A., Astwood, J. D., Engel, K. H., & Jany, K. D. (2006). Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of genetically modified crops. 1. Assessing analytical validation. *Journal of Agricultural Food Chem*, 54(6), 2154-2161, <http://pubs.acs.org/doi/abs/10.1021/jf0523566> , (accessed 12 march 2012).
- [44] Herman, E. (2003). Genetically modified soybeans and food allergies. *Journal of Experimental Botany*, 54(386), 1317-1319, <http://www.ask-force.org/web/Soya/Herman-GM-Soy-Allergies-2003.pdf>, (accessed 12 march 2012).
- [45] EFSA. GMO (2008). Panel Working Group on Animal Feeding Trials. Safety and nutritional assessment of GM plants and derived food and feed: The role of animal feeding trials. *Food and Chemical Toxicology (suppl.)*, 46(1), s2-s70.
- [46] Ocana, M. F., Fraser, P. D., Patel, R. K., Halket, J. M., & Bramley, P. M. (2007). Mass spectrometric detection of CP4EPSPS in genetically modified soya and maize. *Rapid Communication. Mass Spectrometry*, 21(3), 319-328, <http://onlinelibrary.wiley.com/doi/10.1002/rcm.2819/abstract> , (accessed 13 march 2012).
- [47] Boxtel, E.L.V. (2007). Protein quaternary structure and aggregation in relation to allergenicity. *Dissertation, Graduate School VLAG, chapter 1*.
- [48] L'Hocine, L., & Boye, J. I. (2007). Allergenicity of soybean: New developments in identification of allergenic proteins, cross-reactivities and hypoallergization technologies. *Critical Reviews in Food Science and Nutrition*, 47(2), 127-143, <http://>

www.mendeley.com/research/allergenicity-soybean-new-developments-identification-allergenic-proteins-crossreactivities-hypoallergenization-technologies/#, (accessed 13 march 2012).

- [49] Wilson, S., Blaschek, K., & De Mejia, E. G. (2005). Allergenic Proteins in Soybean: Processing and Reduction of 34 Allergenicity. (accessed 10 march 2012). *Nutrition Reviews*, 63(2), 47-58, <http://onlinelibrary.wiley.com/doi/10.1111/j.1753-4887.2005.tb00121.x/abstract>.
- [50] Komatsu, S., Kobayashi, Y., Nishizawa, K., Nanjo, Y., & Furukawa, K. (2010). Comparative proteomics analysis of differentially expressed proteins in soybean cell wall during flooding stress. *Amino Acids*, 39(5), 1435-1449, <http://www.springerlink.com/content/e017718020408605/>, (accessed 12 march 2012).
- [51] Nanjo, Y., Maruyama, K., Yasue, H., Yamaguchi-Shinozaki, K., Shinozaki, K., & Komatsu, S. (2011). Transcriptional responses to flooding stress in roots including hypocotyl of soybean seedlings. *Plant Molecular Biology* <http://www.springerlink.com/content/j1783j7116046564/> (accessed 11 march 2012), 77(1-2), 129-144.
- [52] Shi, F., Yamamoto, R., Shimamura, S., Hiraga, S., Nakayama, N., Nakamura, T., Yukawa, K., Hachinohe, M., Matsumoto, H., & Komatsu, S. (2008). Cytosolic ascorbate peroxidase 2 (cAPX 2) is involved in the soybean response to flooding. *Phytochemistry*, 69, 1295-1303, <http://www.sciencedirect.com/science/article/pii/S0031942208000356>, (accessed 10 mai 2012).
- [53] Toorchi, M., Yukawa, K., Nouri, M.Z., & Komatsu, S. (2009). Proteomics approach for identifying osmotic-stress-related proteins in soybean roots. *Peptides*, 30, 2108-2117, <http://www.sciencedirect.com/science/article/pii/S0196978109003647>, (accessed 09 mai 2012).
- [54] Alam, I., Sharmin, S. A., Kim, K. H., Yang, J. K., Choi, M. S., & Lee, B. H. (2010). Proteome analysis of soybean roots subjected to short-term drought stress. *Plant and Soil*, 333(1-2), 491-505, <http://www.springerlink.com/content/g9813jp3t5687604/>, (accessed 12 april 2012).
- [55] Zhen, Y., Qi, J. L., Wang, S. S., Su, J., Xu, G. H., Zhang, M. S., Miao, L., Peng, X. X., Tian, D., & Yang, Y. H. Comparative proteome analysis of differentially expressed proteins induced by Al toxicity in soybean. *Physiologia Plantarum*, 131(4), 542-554, <http://onlinelibrary.wiley.com/doi/10.1111/j.13993054.2007.00979.x/abstract?systemMessage=Wiley+Online+Library+will+be+disrupted+on+9+June+from+10%3A00-12%3A00+BST+%2805%3A00-07%3A00+EDT%29+for+essential+maintenance&userIsAuthenticated=false&deniedAccessCustomisedMessage=>, (accessed 10 march 2012).
- [56] Duressa, D., Soliman, K., Taylor, R., & Senwo, Z. (2011). Proteomic Analysis of Soybean Roots under Aluminum Stress. *International Journal of Plant Genomics*, doi: 10.1155/2011/282531., <http://www.hindawi.com/journals/ijpg/2011/282531/cta/>, (accessed 10 march 2012).

- [57] Xu, C., Sullivan, J. H., Garrett, W. M., Caperna, T. J., & Natarajan, S. (2008). Impact of solar Ultraviolet-B on the proteome in soybean lines differing in flavonoid contents. *Phytochemistry*, 69(1), 38-48, <http://www.sciencedirect.com/science/article/pii/S0031942207004025>, (accessed 10 march 2012).
- [58] Zhang, Y., Zhao, J., Xiang, Y., Bian, X., Zuo, Q., Shen, Q., Gai, J., & Xing, H. (2011). Proteomics study of changes in soybean lines resistant and sensitive to *Phytophthora sojae*. *Proteome Science*, 9(52), <http://www.proteomesci.com/content/9/1/52>, (accessed 26 february 2012).
- [59] Mithofer, A., Muller, B., Wanner, G., & Eichacker, L. A. (2002). Identification of defence-related cell wall proteins in *Phytophthora sojae*-infected soybean roots by ESI-MS/MS. *Molecular Plant Pathology*, 3(3), 163-166, <http://onlinelibrary.wiley.com/doi/10.1046/j.13643703.2002.00109.x/abstract;jsessionid=436F73B1434802655BD95CC22314118C.d01t04?deniedAccessCustomisedMessage=&userIsAuthenticated=false>, (accessed 26 february 2012).
- [60] Afzal, A. J., Natarajan, A., Saini, N., Iqbal, M. J., Geisler, M., El Shemy, H. A., Mungur, R., & Willmitzer, L. (2009). Lightfoot DA: The nematode resistance allele at the *rhg1* locus alters the proteome and primary metabolism of soybean roots. *Plant Physiology*, 151(3), 1264-1280, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2773059/>, (accessed 10 march 2012).
- [61] Wan, J., Torres, M., Ganapathy, A., Thelen, J., Da, Gue. B., Mooney, B., Xu, D., & Stacey, G. (2005). Proteomic analysis of soybean root hairs after infection by *Bradyrhizobium japonicum*. *Molecular Plant-Microbe Interactions*, 18(5), 458-467, <https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/3478/ProteomicAnalysisSoybeanRootHairs.pdf?sequence=1>, (accessed 27 april 2012).
- [62] Pandurangan, S., Pajak, A., Molnar, S. J., Cober, E. R., Dhaubhadel, S., Hernandez-Sebastia, C., Kaiser, W. M., Nelson, R. L., Huber, S. C., & Marsolais, F. (2012). Relationship between asparagine metabolism and protein concentration in soybean seed. *Journal of Experimental Botany*, 63(8), 3173-3184, <http://jxb.oxfordjournals.org/content/63/8/3173>, (accessed 11 march 2012).
- [63] Sakata, K., Ohyanagi, H., Nobori, H., Nakamura, T., Hashiguchi, A., Nanjo, Y., Mikami, Y., Yunokawa, H., & Komatsu, S. (2009). Soybean Proteome Database: A data resource for plant differential omics. *Journal of Proteome Research*, 8(7), 3539-3548, <http://pubs.acs.org/doi/abs/10.1021/pr900229k>, (accessed 11 march 2012).

IntechOpen

IntechOpen