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

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

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

Soybean Proteomics: Applications and Challenges

Alka Dwevedi and Arvind M Kayastha

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/52408

1. Introduction

Proteomics is one of the most explored areas of research based on global-scale analysis of proteins. It leads to direct understanding of function and regulation of genes. Significant advances in the comprehensive profiling, functional analysis, and regulation of plant proteins have not advanced much as compared to model organisms such as yeast, humans etc. The application of proteomic approaches to plants implicates; comprehensive identification of proteins, their isoforms, as well as their prevalence in each tissue, characterizing the biochemical and cellular functions of each protein and the analysis of protein regulation and its relation to other regulatory networks [1]. Genes of higher eukaryotes (including plants) contain introns which are large and numerous. Therefore, combinational exon usage originating from complex gene structures results in a multitude of splice variants leading to generation of different protein products from a given gene. Thus, the determination of the comprehensive pattern of expression of each protein isoform is a challenging task, most importantly for poorly expressed proteins [2].

The two-dimensional gel electrophoresis (2-DE) is used for profiling protein expression involving separation of complex protein mixtures by molecular charge in the first dimension and by mass in the second dimension. Recent advancement in 2-DE has improved resolution and reproducibility but still automation in high-throughput setting is lagging. The alternative approaches like multi-dimensional protein identification technology involving large-scale proteomics are able to generate a large catalog of proteins present in complex cell extracts. Further, detection of low abundance proteins using sub-cellular fractionation reduces the complexity of protein extracts. These efforts have successfully characterized nuclear, chloroplast, amyloplast, plasma membrane, peroxisome, endoplasmic reticulum, cell wall, and mitochondrial proteomes of a model plant, *Arabidopsis*. Although, high-throughput technologies have helped in characterization of *Arabidopsis* and other organisms' pro-

teomes, characterization of various protein classes including membrane and hydrophobic proteins which are recalcitrant to isolation and analysis is still inaccessible [3].

Food allergy can be a serious nutritional problem in children and adults. Any protein-containing food has the potential to elicit an allergic reaction in the human population. Antibody IgE-mediated reactions are the most prevalent allergic reactions to food. These responses occur after the release of chemical mediators from mast cells and basophils as a result of interactions between food proteins and specific IgE molecules on the surface of these receptor cells. Eight foods or food groups have been identified as the most frequent sources of human food allergens and account for over 90% of the documented food allergies worldwide. These foods are milk, eggs, fish, crustaceans, wheat, peanuts, tree nuts and soy [4]. Despite their well-documented allergenicity, soy derivatives continue to be increasingly used in a variety of food products due to their well-documented health benefits. Soybean has also been one of the selected target crops for genetic modification (GM). For example, the artificial introduction of 5-enolpyruvylshikimate-3-phosphate synthase in soybean crop creates an alternative pathway which is insensitive to glyphosate (most potent herbicide), thus increasing overall crop yield. One of the major concerns regarding the safety of GM foods is the potential allergenicity of the resulting products, namely the possible occurrence of either altered or *de novo* expressed of endogenous allergens after genetic manipulation. This concern justifies careful plant characterization [5]. Proteomics is one of the powerful approaches allowing rapid and reliable protein identification. It can provide information about their post-translational modifications, sub-cellular localization, level of protein expression and protein-protein interactions. Despite the importance of soybean and the availability of powerful tools for the analysis of proteins from sub-cellular organelles, and specifically for the identification of allergens, only a limited number of reports have been published to date.

Soybean is an important source of protein for human and animal nutrition, as well as a major source of vegetable oil. Although soybean is adapted to grow in a range of climatic conditions including adverse environmental and biological factors, still it has been affected with respect to growth, development, and global production For instance, drought reduces the yield of soybean by about 40%, affecting all stages of plant development from germination to flowering thus reducing the quality of the seeds. [6]. Several other abiotic stresses, such as flooding, high temperature, irradiation, or the presence of pollutants in the air and soil have detrimental effects on the growth and productivity of soybean. Along with morphological and physiological studies on the responses of plants to stress conditions, several molecular mechanisms from gene transcription to translation as well as metabolites were investigated. Recent advances in the field of proteomics have created an opportunity for dissecting quantitative traits in a more meaningful way. Proteomics can investigate the molecular mechanisms of plants' responses to stresses and provides a path toward increasing the efficiency of indirect selection for inherited traits. In soybean a comprehensive functional genomics is yet to be performed; therefore, proteomics approaches form a powerful tool for analyzing the functions of complete set of proteins including those involved in stress protection.

2. Proteomics: isolation, identification and classification

In plant proteomics, the type of the plant species, tissues, organs, cell organelles, and the nature of desired proteins affect the techniques that can be used for protein extraction. Furthermore, the extraction process becomes more tedious when the protein is present inside vacuoles, rigid cell walls, or membrane plastids. A perfect protein extraction method involves complete solubilization of total proteins from a given sample and minimizing post-extraction artifact formation, proteolytic degradation as well as removal of non-proteinaceous contaminants. To date, only the proteome of *Arabidopsis* and rice have been studied while less attention has been paid to other plants including soybean. Soybean has high levels of phenolic compounds, proteolytic and oxidative enzymes, terpenes, organic acids, and carbohydrates due to which protein extraction is very tedious. Further it contains contains large quantities of secondary metabolites, *viz.* flavone glycosides (kaempferol and quercetin glycosides), phenolic compounds, lipids and carbohydrates. Thus impedes high-quality protein extraction in turn high-resolution protein separation in 2-DE.

In classical proteome analyses, proteins are initially separated by a 2-DE technique with iso-electric focusing (IEF) as the first dimension and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) as the second dimension. A greater resolution in protein separation has been achieved by introducing immobilized pH gradients (IPGs) for the first dimension. Methodological advances in 2-DE have led to the introduction of two-dimensional fluorescence difference gel electrophoresis (2D-DIGE), which has been used for the comparative analysis of the proteome of soybean subjected to abiotic and biotic stresses [7]. The separated proteins can be subsequently identified by sequencing or by mass spectrometry. By introduction of mass spectrometry into protein chemistry, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and liquid chromatography/tandem mass spectrometry (LC-MS/MS) have become the methods of choice for high-throughput identification of proteins. An alternative technique known variously as 'gel-free proteomics', 'shotgun proteomics', or 'LC-MS/MS-based proteomics' can also be used in high-throughput protein analysis. This approach is based on LC separation of complex peptide mixtures coupled with tandem mass spectrometric analysis. A multidimensional protein identification technology (MudPIT) that usually incorporates separation on a strong cation exchange, reverse-phase column and MS/MS analysis helps the efficient separation of complex peptide mixtures. The gel-free technique have the advantage of being capable of identifying low-abundance proteins, proteins with extreme molecular weights or pI values, and hydrophobic proteins that cannot be identified by using gel-based technique. A combination of gel-based and gel-free proteomics has been used for identification of soybean plasma membrane proteins under abiotic stress, *viz.* flooding, osmotic, salinity stress. Methods for protein identification are not usually organism specific, and they can be applied to a wide range of living organisms in addition to soybean. Identification of proteins is normally performed by using a database search engine such as MASCOT or SEQUEST.

Soybean has an estimated genome size of 1115 Mbp, which is significantly larger than those of other crops, such as rice (490 Mbp) or sorghum (818 Mbp). Sequencing of the 1100 Mbp of

total soybean genome predicts the presence of 46,430 protein-encoding genes, 70% more than in *Arabidopsis* [8]. The soybean genome database contains 75,778 sequences and 25,431,846 residues have been constructed on the basis of the Soybean Genome Project, DOE Joint Genome Institute; this database is available at http://www.phytozome.net. Although the genome sequence information is almost completed, no high-quality genome assembly is available because the results from the computational gene-modeling algorithm are imperfect. In addition, duplications in the genome of soybean result in nearly 75% of the genes being present as multiple copies, which further complicate the analysis. The soybean proteome database (http://proteome.dc.affrc.go.jp/Soybean) provides valuable information including 2-DE maps and functional analysis of soybean proteins. However, the presence of a considerable number of proteins with unknown functions highlights the limitations of bioinformatics prediction tools and the need for further functional analyses. The cellular proteomics helps in identification of changes in protein expression under different growing condition and treatments. The analytical methodology for the separation and identification of a large numbers of proteins should be authentic and confirmable. The proteome map of mature dry soybean seeds has been prepared by employing robotic automation at subsequent steps of 2-DE. Further, UniGene database was implemented for proteins identifications. Total protein from mature dry soybean (*Glycine max* cv. Jefferson) seed was isolated and 2D-PAGE performed using 13 cm IPG strips and subsequently doing SDS-PAGE. Protein spots were analyzed using Phoretix 2D-Advanced software. Excised protein spots were arrayed into 96-well plates and transferred to a Multiprobe II EX liquid handling station for subsequent destaining, tryptic digestion and peptide extraction. MALDI-TOF MS was operated in the positive ion delayed extraction reflector mode. Peptide spectra were submitted to a MS Fit program of Protein Prospector. Assignments from UniGene contigs were subsequently searched against the NCBI non-redundant database using the BLASTP search algorithm to determine similarity matches [9].

Trichloroacetic acid (TCA)/acetone-based and phenol-based buffers are most frequently used in protein extraction from plants. A comprehensive proteomic study was performed on nine organs from soybean plants in various developmental stages by using three different methods for protein extraction and solubilization. The results showed that the use of an alkaline phosphatase buffer followed by TCA/acetone precipitation caused horizontal streaking in 2-DE while use of a Mg/NP-40 buffer followed by extraction with alkaline phenol and methanol/ammonium acetate produced high-quality proteome maps with well-separated spots, high spot intensities, and high numbers of separate protein spots in 2-DE gels [10, 11]. In the case of organelle proteomics particularly that of membrane proteomics, a different extraction procedure is required that involves modifications to dissolve hydrophobic proteins and additional purification steps. Furthermore, when studying protein–protein interactions, it is necessary to extract protein complexes by using buffers with less or no detergent to get the proteins in their native states. Despite the importance of seed filling in the synthesis of storage reserves for germination, systematic proteomic analysis of this phase in legumes is yet to be carried out.

Total seed proteins of soybean (cv. Maverick) at different stages of flowering (14, 21, 28, 35 and 42 days) were isolated and subsequently 2D-PAGE was done. Initially IPG strips of pH 3 to 10 were taken then narrowed down to pH range to 4 to 7 for high-resolution proteome maps. A total of 488 and 679 proteins were identified from 2D-PAGE gels of pH range 4 to 7 and 3 to 10 gels, respectively. Each of the 679 proteins was excised from reference gels for identification by MALDI-TOF MS and a total of 422 proteins (62%) were identified. One unique protein was often represented by more than one spot on the 2D-PAGE gel, most likely due to post-translational modifications or genetic isoforms. Taking into account this redundancy, 216 unique proteins out of 422 were identified. A total of 82 proteins were associated with metabolism (the largest functional class) and the second largest functional class were comprised of 52 spots assigned to the seed storage proteins β-conglycinin and glycinin. An overall down- and up-regulation was observed for metabolism and storage related proteins, respectively, during seed filling, suggesting metabolic activity curtails as seeds approach maturity. Abundance of proteins related to metabolite transporter, disease and defense, energy production, cell growth and division, signal transduction, protein synthesis and secondary metabolism did not vary significantly. Furthermore, 13 sucrose-binding proteins have been mapped to the same UniGene accession number, suggesting the importance of sucrose as a signaling molecule in seed and embryo development. There were a total of 92 unknown proteins which could not be classified, therefore grouped into five expression profiles [12].

3. Implication of proteomics in understanding soybean stress

Soybean is grown worldwide with an average protein content of 40% (highest protein content with respect to other food crops) and oil content of 20% (which is second only to that of groundnut among the leguminous foods). Furthermore, soybean improves soil fertility by fixing nitrogen from the atmosphere in symbiosis with nitrogen fixing bacteria. It is, however, susceptible to various types of stresses (abiotic and biotic). Tolerance and susceptibility to stresses are complex phenomena because they are quantitatively inherited and can occur during different stages of plant growth and development. Extrinsic stress is regarded as the most important stress agent, which results from changes in abiotic factors such as temperature, climatic factors and chemical components, either naturally occurring or manmade. Further, biotic stresses (occurs as a result of damage done to plants by other living organisms, such as bacteria, viruses, fungi, parasites, beneficial and harmful insects, weeds, bacterial, fungal, algal and viral diseases) can also cause huge deterioration in plant growth and yield. Plants have developed adaptive features against these stresses. The genome remains unchanged to a large extent in any particular cell while proteins change dramatically as genes are turned on or off in response to stress. The proteome determines the cellular phenotype and its plasticity in response to external signals. It is proteins that are directly involved in both normal and stress-associated biochemical processes. Therefore, a more complete understanding of stress in soybean may be gained by looking directly into the proteins within a stressed cell or tissue. Proteomic based techniques that allow large-scale protein profiling

are powerful tools for the identification of proteins involved in stress-responses in plants. Extensive studies have evaluated changes in protein levels in plant tissues in response to stresses. Unfortunately, these studies have been mainly focused on non-legume species such as *Arabidopsis* and rice, and only recently have been enlarged to include some legumes. As a result only a handful of studies have been carried out in legumes, although in the next few years there should be a significant increase in the number of legume species and stresses would be analyzed. Recently, proteomic approaches have been applied to various legumes like *M. truncatula*, lentils, lupin, common bean, cowpea and soybean to identify proteins involved in the response to different stresses. Interestingly, many of the induced proteins from these different stresses were common or belonged to overlapping pathways [13].

Considerable amount of research has been carried out during the last decade to find the effect of stress under extreme . These include chloroplast membrane, cell wall and nuclear envelope, while some researchers have focused on individual tissues *viz*. seeds, mitochondria, root tips, vacuoles, chloroplasts and thylakoids. To date, lots of reports have come which emphasize changes in protein expression levels during a particular or integrative stress consequently affecting cellular metabolism. Proteomics provides direct assessment of the biochemical processes of monitoring the actual proteins performing signaling, enzymatic, regulatory and structural functions encoded by the genome and transcriptome.

Following are the different categories of proteins with important properties, which have been shown to play a crucial role against abiotic environmental stress as well as biotic stress. The data so collected from various plants including soybean is based on 2-DE, mass spectrometry and bioinformatics tools.

(a) *Antioxidants Enzymes*

Reactive oxygen species (ROS) in plant cellulars are produced as a consequence of myriad stimuli ranging from abiotic and biotic stress, production of hormonal regulators, as well as cell processes such as polar growth and programmed cell death [14]. These reactive molecules are generated at a number of cellular sites, including mitochondria, chloroplasts, peroxisomes, and at the extracellular side of the plasma membrane. ROS trigger signal transduction events, such as mitogen-activated protein kinase cascades eliciting specific cellular response.s. The influence of these molecules on cellular processes is mediated by both the perpetuation of their production and their amelioration by scavenging enzymes such as superoxide dismutase, ascorbate peroxidase, and catalase. The location, amplitude, and duration of production of these molecules are determined by the specificity of the responses [15]. Accumulation of ROS as a result of various environmental stresses is a major cause of loss of crop productivity worldwide. ROS affect many cellular functions by damaging nucleic acids, oxidizing proteins, and causing lipid peroxidation. It is important to note that whether ROS will act as damaging, protective or signaling factors depends on the delicate equilibrium between ROS production and scavenging at the proper site and time. ROS can damage cells as well as initiate responses such as new gene expression. The cell response evoked is strongly dependent on several factors. The subcellular location for formation of ROS may be especially important for a highly reactive ROS, because it diffuses only a very short distance before reacting with a cellular molecule. Stress-induced ROS accumulation is

counteracted by enzymatic antioxidant systems that include a variety of scavengers, such as superoxide dismutase, ascorbate peroxidase, glutathione peroxidase, glutathione S-transferase, catalase and non-enzymatic low molecular metabolites, such as ascorbate, glutathione (red.), α-tocopherol, carotenoids and flavonoids. In addition, proline can now be added to an elite list of non-enzymatic antioxidants that microbes, animals, and plants need to counteract the inhibitory effects of ROS [16]. Plant stress tolerance may therefore be improved by the enhancement of *in vivo* levels of antioxidant enzymes. The antioxidants as described are found in almost all cellular compartments which signify the importance of ROS detoxification for cellular survival. It has also been shown that ROS influence the expression of a number of genes and signal transduction pathways which suggest that cells have evolved strategies to use ROS as biological stimuli and signals that activate and control various genetic stress-response programs. Control of plant pathogens by genetic engineering has targeted ROS for development of pathogen resistant crop varieties [17]. Antisense technology has been used to reduce the capability to scavenge H_2O_2 in case of model plants like *Arabidopsis thaliana* and *Nicotiana tabacum*. In these plants, antioxidant enzymes like catalase and ascorbate peroxidase are under-expressed and it has been found that they were hyper-responsive to pathogen attack. This further confirms that the ability of plant cells to regulate the efficiency in their ROS-removal strategies is a key point in their resistance against pathogens. The technology is yet to be implemented in case of legumes including soybean as intensive research is going on future prospects of the technology.

(b) *Abscissic acid signaling and related protein*

Abscissic acid (ABA) has been implicated in plant response to environmental stress by interfering at different levels with signaling. Its level increases under stress conditions to trigger metabolic and physiological changes [18]. It has become increasingly clear that the isolated abiotic signaling network is controlled by ABA and the biotic network is controlled by salicylic acid, jasmonic acid and ethylene are interconnected at various levels [19]. The concept of marker genes whose expression is believed to be regulated by individual hormones does not do justice to the nature of the network. The apparent cross-talk in stress-hormone signaling makes it difficult to assign a marker gene or a mutant phenotype to a specific hormone-controlled pathway. The signaling network into which the four stress hormones and other signals feed is apparently designed to allow plants to adapt optimally to specific situations by integrating possibly conflicting information from environmental conditions, biotic stress, and developmental as well as nutritional status. Promoter analyses of ABA/stress-responsive genes revealed that a DNA sequence element consisting of ACGTGGC is important for ABA regulation. For the past several years, researchers have been trying to identify transcription factors that regulate the expression of ABA/stress-responsive genes *via* the consensus element, which is generally known as 'Abscisic Acid Response Element' (ABRE). Many basic leucine zipper class DNA-binding proteins that interact with the element have been reported [20]. Researchers have focused on the small subfamily of *Arabidopsis* basic leucine zipper proteins referred to as ABFs (ABRE-binding factors), whose expression is induced by ABA and by various abiotic stresses (i.e., cold, high salt and drought). ABA is involved in responses to environmental stress such as salinity, and is required by the plant for stress tolerance as found recently on soybean studies. The leaf ABA content in salt-tolerant soybean increased signifi-

cantly under salt stress, while in case of salt sensitive soybean has almost negligible increase in ABA. It is thus possible that ABA enhances salt tolerance in soybean [21].

(c) *GABA-related protein*

γ-Aminobutyric acid (GABA) is a non-protein amino acid that is conserved from bacteria through yeast to vertebrates and was discovered in plants over half-a-century ago. It is mainly metabolized through a short pathway called the GABA shunt, because it bypasses two steps of the tricarboxylic-acid (TCA) cycle. The pathway is composed of three enzymes: the cytosolic and mitochondrial glutamate decarboxylase (GAD), GABA transaminase (GABA-T) and succinic semialdehyde dehydrogenase (SSADH). Although there are differences in the subcellular localization of GABA-shunt enzymes in different organisms have been reported (for e.g. in yeast, SSADH is present inside cytosol) [22]. In an alternative reaction, succinic semialdehyde can be converted to GHB (γ-hydroxybutyric acid) through a GHB dehydrogenase (GHBDH) present in animals and recently identified in plants [23]. Interestingly, research of GABA in vertebrates has focused mainly on its role in the context of plant responses to stress, because of its rapid and dramatic production in response to biotic and abiotic stresses. For example, disruption of the unique SSADH gene in *Arabidopsis* results in plants undergoing necrotic cell death caused by the accumulation of reactive oxygen intermediates (ROIs) when they are exposed to environmental stresses [24]. A recent article reports that a gradient of GABA concentration is essential for the growth and guidance of pollen tubes and suggests that this amino acid plays a role in intercellular signaling in plants, possibly similar to its role in animals. The main question raised by these recent findings is whether GABA itself serves as a signaling molecule in plants. If so, this would imply that GABA is capable of mediating developmental changes and cell guidance by interacting with specialized plant receptors [25].

(d) *Mitogen-activated protein kinase signaling and related proteins*

Like other eukaryotes, plants use mitogen-activated protein kinase (MAPK) cascades to regulate various cellular processes in response to a broad range of biotic and abiotic stress. These cascades promote the transient activation of MAPKs by a dual phosphorylation of Thr and Tyr within the activation loop of the MAPK. Recent studies indicate that MAPKs are not only regulated through phosphorylation by upstream kinases, but also by direct binding of different protein factors [26]. The constitutive activation of MAPKs was found to result in detrimental effects, underlining the importance of a negative regulation of MAPK signaling. MAPK phosphatases (MKPs) are negative regulators of MAPKs. Recent progress in analyzing plant MKP mutants has revealed their important role in fine-tuning MAPK signaling. In particular, the dual-specificity phosphatase MKP1 and the protein tyrosine phosphatase (PTP1) negatively regulate defense responses and resistance to a bacterial pathogen by counter balancing the activation of two MAPKs (MPK3 and MPK6). Interestingly, MKP1 and PTP1 bind CaM, and the phosphatase activity of MKP1 is increased by CaM in a Ca^{2+}-dependent manner. Thus, Ca^{2+} and MAPK signaling pathways appear to be connected through the regulation of plant MAPKs and MKPs by CaM [27].

(e) *Calcium signaling and related proteins*

Plant cells are equipped with highly efficient mechanisms to perceive, transduce and respond to a wide variety of internal and external signals during their growth and development. Perception of signals *via* receptors results in generation or synthesis of non-proteinaceous molecules which are termed as messengers. The messengers include Ca^{2+} ions, small organic molecules such as cyclic nucleotide monophosphates, inositol triphosphates and inorganic molecules such as H_2O_2 and NO. The elements of receptors, messengers, sensors and targets vary depending on the signal received. Identification and functional assignment of these elements in a stimulus-specific signal transduction pathway is a challenging area for plant biologists. With the completion of genome sequences of various organisms, including *Arabidopsis thaliana*, *Oryza sativa*, *Medicago trunculata*, *Glycine max* etc. it has become evident that plants have a large number of motifs containing helix-loop-helix which binds to Ca^{2+} [28]. Further, Ca^{2+} has been implicated in mediating various developmental processes (pollen tube growth, root-hair and lateral root development and nodulation), hormone regulated cellular activities (cell division and elongation, stomatal closure/opening), pathogen- and elicitor-induced defense related processes, and a variety of abiotic stress signal induced gene expression. However, the identity and functions of downstream transducers and mechanisms by which Ca^{2+} mediates a variety of cellular responses are just begin to unravel in plants. In plants, spatially and temporally distinct changes in cellular Ca^{2+} concentrations, designated as "Ca^{2+} signatures" that are evoked in response to different stimuli like drought, salt or osmotic stresses, temperature, light and plant hormones represent a central mechanistic principle to present defined stimulus-specific information [29]. These specific "Ca^{2+} signatures" are formed by the tightly regulated activities of channels and transporters at different membranes and cell organelles. While the identity and function of components of the Ca^{2+} extrusion system are rather well understood in plant cells, the molecular identity of Ca^{2+} specific influx channels has remained unknown. However, non-specific influx of Ca^{2+} mediated by ligand gated cation channels like cyclic nucleotide gated channels and glutamate receptor-like proteins contribute to different Ca^{2+} mediated cellular functions like the response to pathogens, pollen tube growth and abiotic stress. The unique structural composition of Ca^{2+} binding proteins and the complexity of the target proteins regulated by the Ca^{2+} sensors allow the plant to tightly control the appropriate adaptation to its ever changing environment. It is actually still not well understood about interface of information presentation by a specific Ca^{2+} signal and initiation of information decoding by Ca^{2+} sensors that represent a most critical step in specific information processing [30].

4. Significance of proteomics in soybean allergenicity

Soybeans have played a central role in concerns about GM introduced allergens and in using GM to remove intrinsic allergens. Soybean is a rich and inexpensive source of proteins for humans and animals. Soybean milk and dairy product replacement is growing in acceptance, not only by people sensitive to lactose and/or milk proteins, but also for health considerations. Soybean protein is widely used in thousands of processed foods throughout the industrialized world and is a staple crop in Asia. Soybean ranks among the eight most significant food allergens. Soybean sensitivity is estimated to occur in 5-8% of children and

1-2% of adults. The allergic reaction is only rarely life-threatening with the primary adverse reactions to consumption being atopic (skin) reactions and gastric distress. Symptoms of soy allergy usually appear within a few minutes to two hours of eating soy ingredients. People with soy allergies may cross-react with peanuts or other legumes, such as beans or peas. Soy is one of the most common allergens for infants who have not yet begun eating solid foods, because they may be fed soy-based infant formula. It is rare for babies to have a traditional IgE mediated food allergy to soy, but some babies may develop milk-soy protein intolerance [31-34] or food protein induced enterocolitis syndrome [http://foodallergies.about.com/od/ soyallergies/a/Soy-Allergy-Overview.htm]. Infants will usually develop these sensitivities within a few months of birth, and most will outgrow them by the age of two. Most people with soy allergies can tolerate the small amount of soy protein that remains in refined soy-bean oil and soy lecithin. Both of these ingredients may cause allergic reactions in highly sensitized people. There are some data available that describe the natural variation in allergen proteins that occur in soybean. For a better understanding of the variation of allergen proteins that might be expected to occur in GM soybeans, it is important to determine the natural variation of protein composition both in wild and GM soybeans. "Proteomics" approach is the foremost one which allows protein identification and quantification with utmost accuracy.

Biotechnology critics have claimed that an apparent rise in the number of soybean allergic individuals in the UK is correlated with the development of GM soybeans in the American market. GM-soybeans that have been developed in the US include herbicide-resistance (glyphosate) and seeds with higher percentage of essential amino acids, *esp*. methionine. Experiments have directly tested the allergenicity of herbicide-tolerant soybeans using immunological tests with samples from soybean-sensitive people. These assays have shown that herbicide-resistant GM soybeans do not present any measurable differences in allergenicity compared with non-GM soybeans and are, therefore, substantially equivalent by allergenic criteria. Sensitive people remain allergic to GM soybeans, but there is no additional allergenic risk to others. According to some reports protein expressed corresponding to transgene responsible for herbicide-resistance in soybeans has allergenic motifs [35]. On ingestion a portion of the transgene along with the promoter get transferred to human gut bacteria. The transformed bacteria containing transgene continues to produce herbicide-resistance allergenic protein even when the individual is not eating GM soy. Therefore an individual is constantly exposed to potentially allergenic protein, being created within his gut. Further, herbicide-resistant protein is made more allergenic due to its misfolding brought by rearrangement of unstable transgenes. Some reports emphasize the fact that protein allergenicity is due to suppression of pancreatic-enzymes due to which protein remains in the gut for longer duration contributing to allergies. There is insufficient data to support *in vivo* toxicity of herbicide-resistant protein either due to transformation or enzyme suppression [36]. GM-soybeans with enhanced methionine content such as prolamines and 2S albumins were tested for its allergenicity before its commercialization. It was found that allergenicity was much higher with respect to wild soybeans [37]. Consequently the development of GM soybean with enhanced methionine has been abandoned and no product was released, thus nobody was harmed by its adverse reactions. Recently, one of the interesting analyses has been

done on GM-soy irrespective of herbicide resistance or enhanced methionine content. It has found that GM transformation process may lead to increment in natural allergens in soybeans. The level of one known allergen is trypsin inhibitor which is 27% higher in raw GM soy varieties with respect to natural varieties [38]. Further, it has also been found that cooked GM soy has sevenfold higher amount of trypsin inhibitor as compared to cooked non-GM soy due to its extreme heat stability. There are several reports including both supportive as well unsupportive towards effects of GM-soy on humankind as well as on other flora and fauna of the environment. It will require intensive research including proteomics before their release into the commercial markets.

Plant biotechnology has not only tried to produce GM-soy which is herbicide resistance or with enhanced methionine content but also aimed to remove naturally occurring allergens in native soy varieties. Presently primary treatment for food allergies is avoidance, but it is unavoidable in case of soybean protein which is present in thousands of products. Therefore, it is very difficult to avoid soybean and its derived products. Research is going on to produce hypoallergenic variants of soybean which has potential to reduce the risk of adverse reactions. Soybeans possess as many as 15 proteins recognized by IgEs from soybean-sensitive people [39]. The immunodominant soybean allergens are the β-subunit of conglycinin and P34 or Gly m Bd 30k (cysteine proteases from papain family). The P34/Gly m Bd 30k protein is a unique member of the papain superfamily lacking the catalytic cysteine residue that is replaced by a glycine which is 70% more allergenic with respect to conglycinin. There are several approaches that have been taken to produce a hypoallergenic soybean. One approach was to search cultivars which lack allergens and then crossing its germplasm to elite germplasm. This approach could not be implemented as there was no soybean cultivar (either domesticated or wild) present which lack P34/Gly m Bd 30k. Immunological assays of P34/Gly m Bd 30k with antibodies from soybean-sensitive people resulted in the identification of 14 contiguous and non-contiguous linear epitopes. The presence of so many distinct linear epitopes means that the probability of a naturally occurring variant with a sufficient number of alterations to disrupt the allergenicity is extremely small. Protein engineering could be performed to alter amino acid sequence by disrupting allergenic sequences. Using linear peptides to test possible modifications, it is straightforward to assay numerous variants and pick one that is not recognized by the IgE population. The epitope modification approach is not feasible to produce an essentially hypoallergenic variant. The problem with this technology is to remove completely the intrinsic allergen and substitute the `hypoallergenic' variant in its place. Further, the modification of the protein to remove the allergenic epitopes may alter the protein's folding, that, in turn, may affect the protein's intracellular targeting, stability and accumulation. All these possibilities will need to be tested for experimentally and, finally; the newly produced hypoallergenic variant will need to be tested to ensure that it too is not a new allergen. For these reasons, substituting a hypoallergenic variant of a plant still has a high technological threshold and has yet to be achieved. The alternative GM approach is to eliminate the allergen by suppression. There have been several attempts to reduce and/or eliminate allergens using gene suppression technology. Gene-silencing techniques involve transgenic soybeans with eliminated immunodominant human allergen P34/Gly m Bd 30k. It involves complete elimination of the P34/Gly m Bd 30k

allergen from the initial somatic embryos through the third generation homozygous soybeans. Suppression of the allergen did not introduce any changes in the pattern of growth and development of the plant or seed at both the gross and subcellular level. In order to compare the P34-suppressed soybeans with the wild type, large-scale proteomic analysis was performed. Imaging of the 2D gels identified over 1400 individual elements. Mass spectrometry analysis of about 140 of these spots confirmed that the only overt changes in composition in the transgenic soybeans was the suppression of the P34/Gly m Bd 30k protein with no other proteins induced or suppressed [40]. Further analysis with sera samples from soybean-sensitive people confirmed a loss of the P34 allergen and no induction of any new allergens. The proteome and immunological analysis together confirms that it is feasible to suppress an endogenous allergen without introducing adverse effects on the plant or changing the composition of the soybean seed in any way other than the removal of the targeted protein. This result meets the test of `substantial equivalence' where the GM soybean seed is essentially identical except for the change in the single desired characteristic. Suppressing P34/Gly m Bd 30k in GM soybeans is a first step and a demonstration in addressing the growing concerns about food allergies and its relationship to the development of GM crops. More detailed studies and approaches should provide the tests needed to gain regulatory approval in nations that are currently cautious about this technology. Natarajan *et al.* [41] have compared the profiles of allergen and anti-nutritional proteins both in wild and GM soybean seeds. 2D-PAGE was used for the separation of proteins at two different pH ranges and applied a combined MALDI-TOF-MS and LC-MS analysis for the identification of proteins. Although overall distribution patterns of the allergen and anti-nutritional proteins Gly m Bd 60K (conglycinin), Gly m Bd 30K, Gly m Bd 28K, trypsin inhibitors, and lectin appeared similar, there was remarkable variation in the number and intensity of the protein spots between wild and GM soybean. The wild soybean showed fifteen polypeptides of Gly m Bd 60K and three polypeptides of trypsin inhibitors. GM soybean showed twelve polypeptides of Gly m Bd 60K and two polypeptides of trypsin inhibitors. In contrast, the GM soybean showed two polypeptides of Gly m Bd 30K and three polypeptides of lectin and the wild type showed two and one polypeptides of Gly m Bd 30K and lectin, respectively. The same number of Gly m Bd 28K spots was observed in both wild and GM soybean [41].

The fear of allergic reactions has produced much of the concern about the risks of GM crops. In order to broadly apply genetic modification to crops, there is an urgent need for better biochemical and molecular methods, including animal models, to test for food allergens experimentally so that the supporting data can be provided to evaluate newly proposed and actual GM products. In order to design transgenes, it would be useful to predict allergenicity but, currently, there are no models that would permit accurate assessment of allergenic potential of proteins unrelated to known allergens. Liver represents a suitable model for monitoring the effects of a diet, due to its key role in controlling the whole metabolism. Previous studies on hepatocytes from young female mice fed on GM soybean demonstrated nuclear modifications involving transcription and splicing pathways [42, 43]. The morphofunctional characteristics of the liver of 24-month-old mice, fed from weaning on control or GM soybean, were investigated by combining a proteomic approach with ultrastructural, morphometrical and immunoelectron microscopical analyses. Several proteins belonging to

hepatocyte metabolism, stress response, calcium signaling and mitochondria were differentially expressed in GM-fed mice, indicating a more marked expression of senescence markers in comparison to controls. Moreover, hepatocytes of GM-fed mice showed mitochondrial and nuclear modifications indicative of reduced metabolic rate. This study demonstrates that GM soybean intake can influence some liver features, although the mechanisms remain unknown. Therefore, it is required to investigate the long-term consequences of GM-diets, further studies are required for potential synergistic effects with other factors like ageing, stress etc.

5. Challenges and perspectives

Soybean is a species of great agronomic and economic interest. It is one of the most recalcitrant plant species to be used as experimental material in proteomic analysis. Furthermore, there are several difficulties in the study of proteins (irrespective of source) with respect to DNA and RNA. The foremost important thing is the maintenance of secondary and tertiary structure during their analysis. They have problems with easy denaturation on exposure to high temperature, extremes of pH, oxidation, specific chemicals etc. There are some classes of proteins which are difficult to analyze due to their poor solubility. Proteins cannot be amplified like DNA, therefore less abundant species are very difficult to detect. However, many potentially important proteins (in scarce) are lost due to non-specific binding or the co-removal of proteins/peptides intrinsically bound to the high abundant carrier proteins. Following are two methods developed recently to resolve detection of less abundant plant proteins [44]:

- The use of equalizer beads coupled with a combinational library of ligands containing diverse population of beads with equivalent binding capacity to most of the proteins present in a sample.

- The ultra-microarrays have been found to have high specificity and sensitivity with detection levels in the range of attomole (10^{-18} mole).

The current depth of knowledge regarding the soybean proteome is significantly less than that for some other plants. The soybean proteome map which is available in the database (http://proteome.dc.affrc.go.jp/soybean/) corresponds to various types of stresses, allergenicity, and studies on natural product biosynthesis in soybean. The other challenges in plant proteomics including soybean are standardization of methodologies, dissemination of proteomics data into publicly available databases and most importantly its cost expensiveness. Furthermore, most proteomics technologies use complex instrumentation and critical computing power. Currently, there is no expertise available for functional interpretation of data obtained from integration of proteomics with genomics and metabolomics.

The significance of proteomics over genomics and transcriptomics has been debated since the field has emerged. The importance of the proteome cannot be overstated as it is the proteins within the cell that provide structure, produce energy, as well as allow communica-

tion, movement, and reproduction. Basically, proteins provide structural and functional framework for cellular life. Genetic information is static while the protein complement of a cell is dynamic. Differential proteomics is a scientific discipline that detects the proteins associated with a diseased state (either due to abiotic or biotic stress, toxicity due to allergenicity, genetic modifications etc.) by means of their altered levels of expression between the control and diseased states. Extensive research towards the development of a soybean proteome map would permit the rapid comparison of soybean cultivars, mutants, and transgenic lines. Moreover, studies of soybean physiology will also benefit from the existence of a detailed and quantitative proteome reference map of the soybean plant. The information obtained from soybean proteomics will be helpful in predicting the function of plant proteins and will aid in molecular cloning of the corresponding genes in the future. The identification of novel genes, the determination of their expression patterns in response to stress, and an understanding of their functions in stress adaptation will provide us with the basis for effective strategies for engineering improved stress tolerance in soybean. With the advancement of new technologies in proteomics combined with advanced bioinformatics, we are currently identifying molecular signatures of diseases based on protein pathways and signaling cascades. Applying these findings will improve our understanding of the roles of individual proteins or the entire cellular pathways in the initiation and development of disease. The abundance of information provided by proteomics research is entirely complementary with the genetic information being generated by genomics research. Proteomics makes a key contribution to the development of functional genomics. The combination of genomics and proteomics will play a major role in understanding molecular mechanisms in plant pathology, and it will have a significant impact on the development of high yield varieties, with better resistance towards adverse environmental factors as well as various pathogenic diseases caused by bacteria, viruses and fungi in the future.

Author details

Alka Dwevedi[1*] and Arvind M Kayastha[2*]

*Address all correspondence to: kayasthabhu@gmail.com

1 Regional Centre for Biotechnology, India

2 School of Biotechnology, Faculty of Science, Banaras Hindu University, India

References

[1] Wu, D. D., Hu, X., Park, E., Wang, X., Feng, J., & Wu, X. (2010). Exploratory analysis of protein translation regulatory networks using hierarchical random graphs. *BMC Bioinformatics*, 11(3), S2.

[2] Ezkurdia, I., del Pozo, A., Frankish, A., Rodriguez, J. M., Harrow, J., Ashman, K., Valencia, A., & Tress, M. L. (2012). Comparative proteomics reveals a significant bias toward alternative protein isoforms with conserved structure and function. *Molecular Biology and Evolution*, 29-2265.

[3] Bertone, P., & Snyder, M. (2005). Prospects and challenges in proteomics. *Plant Physiology*, 138-560.

[4] Hefle, S. L, Nordlee, J. A, & Taylor, S. L. (1996). Allergenic foods. *Critical Reviews in Food Science and Nutrition*, 36, S69-S89.

[5] Metcalfe, D. D. (2005). Genetically modified crops and allergenicity. *Nature Immunology*, 6-857.

[6] Manavalan, L. P, Guttikonda, S. K, Tran, L. S, & Nguyen, H. T. (2009). Physiological and molecular approaches to improve drought resistance in soybean. Plant & Cell Physiology , 50-1260.

[7] Atkinson, N. J, & Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. *Journal of Experimental Botany*, 63-3523.

[8] Nouri-Z, M., Toorchi, M., & Komatsu, S. (2011). Chapter 9: Proteomics approach for identifying abiotic stress responsive proteins in soybean. *In Soybean-molecular aspects of breeding ed. Aleksandra Sudaric. Intech, April 11*, 978-9-53307-240-1.

[9] Mooney, B. P., Krishnan, H. B., & Thelen, J. J. (2004). High-throughput peptide mass fingerprinting of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. *Phytochemistry*, 65-1733.

[10] 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*, 342-214.

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

[12] Lei, Z, Elmer, A. M, Watson, B. S, Dixon, R. A, Mendes, P. J., & Sumner, L. W. (2005). A two-dimensional electrophoresis proteomic reference map and systematic identification of 1367 protein from a cell suspension culture of the model legume Medicago truncatula. *Molecular and Cellular Proteomics*, 4, 1812-1825.

[13] Nasi, A., Picariello, G., & Ferranti, P. (2009). Proteomic approaches to study structure, functions and toxicity of legume seeds lectins. Perspectives for the assessment of food quality and safety. *Journal of Proteomics*, 72-527.

[14] Bailey-Serres, J., & Mittler, R. (2006). The roles of reactive oxygen species in plant cells. *Plant Physiology*, 141, 311.

[15] Mittler, R. (2002). Oxidative stress, antioxidants and stress tolerance. *Trends in Plant Science*, 7-405.

[16] Daft, J., Vandenabeele, S., Vranová, E., Van Montagu, M., Inzé, D., & Van Breusegem, F. (2000). Dual action of the active oxygen species during plant stress responses. *Cellular and Molecular Life Sciences*, 57-779.

[17] Collinge, D. B., Jørgensen, H. J .L., Lund, O. S., & Lyngkjær, M. F. (2010). Engineering pathogen resistance in crop plants: Current Trends and Future Prospects. *Annual Review of Phytopathology*, 48-269.

[18] Leung, J., & Giraudat, J. (1998). Abscisic acid signal transduction. *Annual Review of Plant Physiology and Plant Molecular Biology*, 49-199.

[19] Mauch-Mani, B., & Mauch, F. (2005). The role of abscisic acid in plant-pathogen interactions. *Current Opinion in Plant Biology*, 8-409.

[20] Choi, H. I., Hong, J. H., Ha, J. O., Kang, J. Y., & Kim, S. Y. (2000). ABFs, a family of ABA-responsive element binding factors. *The Journal of Biological Chemistry*, 275-1723.

[21] Xu , Y. X., Fan, R., Zheng, R., Li, M. C., & Yu, Y. D. (2011). Proteomic analysis of seed germination under salt stress in soybeans. *Journal of Zhejiang University-Science B (Biomedicine & Biotechnology)*, 12-507.

[22] Bouche´, N., Lacombe, B., & Fromm, H. (2003). GABA signaling: a conserved and ubiquitous mechanism. *Trends in Cell Biology*, 13-607.

[23] Roberts, M. R. (2007). Does GABA Act as a Signal in Plants? Plant Signaling & Behavior , 2-408.

[24] Bouché, N., Fait, A., Bouchez, D., Møller, S. G., & Fromm, H. (2003). Mitochondrial succinic-semialdehyde dehydrogenase of the γ-aminobutyrate shunt is required to restrict levels of reactive oxygen intermediates in plants. *Proceedings of National Academy of Sciences of United States of America*, 100-6843.

[25] Ludewig, F., Hüser, A., Fromm, H., Beauclair, L., & Bouché, N. (2008). Mutants of GABA transaminase (POP2) suppress the severe phenotype of succinic semialdehyde dehydrogenase (SSADH) mutants in Arabidopsis. Plos One , 3, e3383.

[26] Wrzaczek, M., & Hirt, H. (2001). Plant MAP kinase pathways: how many and what for? *Biology of the Cell*, 93-81.

[27] Zhang, T., Liu, Y., Yang, T., Zhang, L., Xu, L., & An, L. (2006). Diverse signals converge at MAPK cascades in plants. *Plant Physiology and Biochemistry*, 44-274.

[28] Yang, T., & Poovaiah, B. W. (2003). Calcium/calmodulin-mediated signal network in plants. *Trends in Plant Science*, 8-505.

[29] Reddy, V. S., & Reddy, A. S. N. (2004). Proteomics of calcium-signaling components in plants. *Phytochemistry*, 65-1745.

[30] Perochon, A, Aldon, D, Galaud, J. P., & Ranty, B. (2011). Calmodulin and calmodulin-like proteins in plant calcium signaling. *Biochimie*, 93, 2048-2053.

[31] Dwevedi, A., & Kayastha, A. M. (2010). Plant β-galactosidases: Physiological significance and recent advances in technological applications. *Journal of Plant Biochemistry and Biotechnology*, 19-9.

[32] Dwevedi, A., & Kayastha, A. M. (2011). Chapter 11: Soybean, a multifaceted legume with enormous economic capability. *In Soybean: Biochemistry, Chemistry and Physiology, ed. Tzi-Bun Ng. Intech, April 26*, 978-9-53307-219-7.

[33] Dwevedi, A., & Kayastha, A. M. (2009). A β-galactosidase from pea seeds (PsBGAL): purification, stabilization, catalytic energetics, conformational heterogeneity, and its significance. *Journal of Agricultural and Food Chemistry*, 57-7086.

[34] Kishore, D., Talat, M., Srivastava, O. N., & Kayastha, A. M. (2012). Immobilization of β-galactosidase onto functionalized graphene nano-sheets using response surface methodology and its analytical applications towards milk and whey lactose. *Plos One*, 7, e40708.

[35] Burks, A.W., & Fuchs, R. L. (1995). Assessment of the endogenous allergens in glyphosate tolerant and commercial soybean varieties. *Journal of Allergy and Clinical Immunology*, 96-1008.

[36] Mazur, B. J, & Falco, S. C. (1989). The development of herbicide resistant crops. *Annual Review of Plant Physiology and Plant Molecular Biology*, 40-441.

[37] Herman, E. M. (2003). Genetically modified soybeans and food allergies. *Journal of Experimental Botany*, 54-1317.

[38] Domingo, J. L. (2007). Toxicity studies of genetically modified plants. *Critical Reviews in Food Science and Nutrition*, 47721-733.

[39] Burks, A. W., Brooks, J. R, & Sampson, H. A. (1988). Allergenicity of major component proteins of soybean determined by enzyme-linked immunosorbent assay (ELISA) and immunoblotting in children with atopic dermatitis and positive soy challenges. *Journal of Allergy and Clinical Immunology*, 81-1135.

[40] Helm, R. M., Cockrell, G., Herman, E., Burks, A. W., Sampson, H. A., & Bannon, G. A. (1998). Cellular and molecular characterization of a major soybean allergen. *International Archive of Allergy Immunology*, 117-29.

[41] Natarajan, S. S., Xu, C., Bae, H., Caperna, T. J., & Garrett, W. M. (2006). Proteomic analysis of allergen and antinutritional proteins in wild and cultivated soybean. *Journal of Plant Biochemistry and Biotechnology*, 15-103.

[42] Malatesta, M., Caporaloni, C., Gavaudan, S., Rocchi, M. B., Serafini, S., Tiberi, C., & Gazzanelli, G. (2002). Ultrastructural morphometrical and immunocytochemical analyses of hepatocyte nuclei from mice fed on genetically modified soybean. *Cell Structure and Function*, 27-173.

[43] Malatesta, M., Boraldi, F., Annovi, G., Baldelli, B., Battistelli, S., Biggiogera, M., & Quaglino, D. (2008). A long-term study on female mice fed on a genetically modified soybean: effects on liver ageing. *Histochemistry and Cell Biology*, 130-967.

[44] Cho, W. C. S. (2007). Proteomics technologies and challenges. *Genomics, Proteomics & Bioinformatics*, 5-77.