



Separomics applied to the proteomics and peptidomics of low-abundance proteins: Choice of methods and challenges – A review

Maria Cristina Baracat-Pereira¹, Meire de Oliveira Barbosa¹, Marcos Jorge Magalhães Júnior¹, Lanna Clícia Carrijo¹, Patrícia Dias Games¹, Hebréia Oliveira Almeida¹, José Fabiano Sena Netto¹, Matheus Rodrigues Pereira² and Everaldo Gonçalves de Barros²

¹*Departamento de Bioquímica e Biologia Molecular, Instituto de Biotecnologia Aplicada à Agropecuária, Universidade Federal de Viçosa, Viçosa, MG, Brazil.*

²*Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, MG, Brazil.*

Abstract

The enrichment and isolation of proteins are considered limiting steps in proteomic studies. Identification of proteins whose expression is transient, those that are of low-abundance, and of natural peptides not described in databases, is still a great challenge. Plant extracts are in general complex, and contaminants interfere with the identification of proteins involved in important physiological processes, such as plant defense against pathogens. This review discusses the challenges and strategies of separomics applied to the identification of low-abundance proteins and peptides in plants, especially in plants challenged by pathogens. Separomics is described as a group of methodological strategies for the separation of protein molecules for proteomics. Several tools have been used to remove highly abundant proteins from samples and also non-protein contaminants. The use of chromatographic techniques, the partition of the proteome into subproteomes, and an effort to isolate proteins in their native form have allowed the isolation and identification of rare proteins involved in different processes.

Key words: sample preparation, complex protein extract, subproteomes, low-abundance proteins, cell wall proteins.

Introduction

The separomics challenges in plants

Proteomics tools have been widely used in recent years. Proteomics and peptidomics involve sophisticated methodologies which accurately detect alterations respectively in protein and peptide synthesis, under different physiological situations. Two main tools are widely used to isolate proteins, especially so two-dimensional electrophoresis (2-DE) associated with mass spectrometry (MS), and liquid chromatography associated with MS (LC-MS). Yet both present limitations inherent to the techniques (Cho, 2007). Multi-dimensional liquid chromatography has been valued in recent years as a technique to obtain native samples, due to the need to validate biological events observed in proteomic studies. Obtaining native proteins is one of the biggest challenges in proteomics. Difficulties in isolating and identifying protein and peptide groups occur due to the high complexity of proteins present in the samples, the presence of non-protein contaminants that are difficult to

remove and the occurrence of post-translational modifications. The difficulties are even greater for proteins and peptides present in low abundance (LAP) in the tissues. These, however, have attracted the attention of researchers since they are in general very effective and/or transiently present as components of finely controlled metabolic pathways. Thus, alternative methods to detect and identify these proteins are necessary.

Alternative strategies applied to the extraction, purification and biochemical and functional analyses of these molecules have been proposed, favoring access to structural and functional information of hard-to-reach proteins and peptides (Kolodziejek and van der Hoorn, 2010). Separomics is described as a group of methodological strategies aimed at separating protein molecules for proteomics, including fractionation and enrichment of specific molecules (Fang and Zhang, 2008). The use of separomic tools is especially important for peptidomics, which is described as the group of methodologies and procedures applied to the analysis of native peptides by means of proteomics tools, since they are small and non-abundant proteins (Jurgens and Schrader, 2002).

Especially for peptidomics, 2-DE is difficult to apply, due to the low concentration of the peptide molecules, their

small sizes (up to 10 kDa), their partial hydrophobic character, and their ionic characteristics, as many peptides are strongly cationic. For the identification of these molecules, the greatest challenges are the small number of available specific databases and the low number of studied and posted molecules, which makes their identification through limited proteolysis techniques and MALDI-MS difficult. In addition, the partial hydrophobicity characteristics and surface charges facilitate peptide molecular associations, making them unavailable for analysis by any known proteomics tools.

For plant proteomics, the greatest challenges are to reduce sample complexity and remove contaminants which are incompatible with the isolation and identification tools (Kolodziejek and van der Hoorn, 2010). The correct use of separomics is imperative, especially for the identification of proteins and peptides expressed in low concentrations. This said, the current review aimed at discussing difficulties and challenges in plant proteomics and peptidomics, and to point out methodologies and strategic tools capable of detecting low-abundance proteins, especially biological peptides, differentially expressed in soybean plants after biotic and abiotic stresses.

Proteomic analysis applied to the response of soybean plants to pathogens

Proteomic analysis has become the most powerful tool for the functional characterization of plants. Information on soybean sequencing (Kim *et al.*, 2010; Schmutz *et al.*, 2010) and the soybean genome database (Phytozome v7.0: *Glycine max*) are available. Soybean was the first legume species to have its complete genome sequenced, becoming therefore a key reference for the more than 20,000 legume species (Schmutz *et al.*, 2010). Several genomes are emerging as model for plants, including that from soybeans. Komatsu and Ahsan (2009) have discussed the advantages and limitations of different proteomics tools applied to the study of soybean defense. In the last two years, different research groups have also discussed the difficulties and alternatives concerning proteomic analyses of plants in general (Sudaric *et al.*, 2010; Yamaguchi and Sharp, 2010; Bindschedler and Cramer, 2011). Low protein concentration, difficulties in protein extraction, genome ploidy, interference of highly abundant proteins in green tissue are some of the main challenges in plant proteomics (Bindschedler and Cramer, 2011). In general, there are many additional challenges for plant proteomics, such as the identification of proteins expressed transiently and in relatively low concentration, or the fact that the natural peptide has not yet been deposited in any of the databases. *In silico* information is essential for proteomic analysis of important physiological processes, such as plant defense. The subcellular proteome analysis of soybean plants submitted to stress conditions aims to identify proteins and peptides that are differentially expressed and potentially involved in

plant defense or pathogen resistance induction. Biotechnological methods may then be developed in order to intercept the pathogen action before infection, or to produce defense agents to boost the plant's defense system. Organelle proteins and specific soybean tissues have been studied, such as membrane (Komatsu and Ahsan, 2009; Bindschedler and Cramer, 2011), primary roots (Nouri and Komatsu, 2010) and cell wall (Yamaguchi and Sharp, 2010).

Searching For Native Low-Abundance Proteins and Peptides By Proteomics and Peptidomics

Techniques with different sensitivities and accuracies should be used in the analysis of the proteome. Due to the physical, chemical and biological diversity of proteins, proteomics tools present limitations that make it unfeasible to analyze the entire proteome with only a single separation strategy, even if it is orthogonal, such as the 2-DE or multi-dimensional liquid chromatography. Furthermore, the cellular protein concentration may vary from mg mL⁻¹ to pg mL⁻¹ (Fang and Zhang, 2008; Jorrín-Novo *et al.*, 2009).

Different proteomic analysis platforms were developed in recent decades based on biochemical tools already available for protein isolation and identification. By means of these strategies, thousands of proteins have been identified, especially high-abundance proteins (HAP), whereas the identification of low-abundance proteins (LAP) is still limited. The search for LAP is growing, for they can represent important biological markers (Lescuyer *et al.*, 2007), be responsible for eliciting important cellular responses, and may even correspond to low molecular weight transcription factors. Not only is their synthesis transient but they can also be easily lost (Corthals *et al.*, 2000). Different procedures have been used to reduce HAP levels in extracts and to improve LAP detection, such as: (1) partial removal of RUBISCO by increasing dithiothreitol concentration in extracts of rice leaves (Cho *et al.*, 2008), or by fractionating extracts of soybean leaves with calcium chloride and sodium phytate (Krishnan and Natarajan, 2009), (2) addition of solvents such as the isopropanol to remove storage proteins from extracts of soybean seeds (Natarajan *et al.*, 2009), (3) division of the proteome into subproteomes, as done for banana leaf membranes (Vertommen *et al.*, 2011), or (4) the use of affinity and immunoaffinity columns for protein purification (Azarkan *et al.*, 2007; Fang and Zhang, 2008).

Due to the need for improvement and/or development of protein separation methods for proteomic studies, separomics makes its appearance as the science for proteomic separation. The main goal of separomics is to obtain a specific set of proteins in a given biological system for proteome composition analysis, for protein-protein interaction studies, or for analysis of alterations in protein synthesis in biological materials submitted to different physiological conditions. The concept of *Separomics* (originally called *Seppromics*) was proposed to describe and define technolo-

gies, processes, requirements, patterns and applications on proteomic separation, including fractionation and enrichment (Huang *et al.*, 2005). The greatest challenge was to reduce the complexity of a given proteome, aiming to increase the effectiveness of proteomic analysis as well as the identification of new proteins through mass spectrometry (Fang and Zhang, 2008; Kosová *et al.*, 2011).

Specific structures in plants, such as the cell wall and vacuoles, contain substances responsible for the inferior and non-reproductive outcomes during the separation of proteins due to proteolytic breakdown, streaking and charge heterogeneity. In most plant tissues, the proteins are part of complex structures, requiring special care for their isolation in a soluble form, as a native or non-native protein. The most commonly found interfering molecules during protein isolation are phenolic compounds, proteolytic and oxidative enzymes, terpenes, pigments, organic acids, inhibiting ions, and carbohydrates (Carpentier *et al.*, 2005).

A further limiting factor in proteomic studies in plants is the loss of protein solubility caused by either the enrichment of protein extracts or by the properties of the solvents. Solubility of proteins is usually dependent on their concentration and amino acid composition, type and solvent dielectric constant, ionic strength, and on the presence of contaminants (Jorrín-Novo *et al.*, 2009). The optimum solubility conditions will be empirically defined for each system.

Preparing and Isolating Low-Abundance Proteins and Peptides From Plants

Plant extracts require specific sample preparation procedures especially adjusted for proteomic and peptidomic studies. These procedures should take into account the solubility, the physico-chemical characteristics and the cellular localization of the proteins, as well as the presence of interfering molecules (Chen and Harmon, 2006; Matros *et al.*, 2010), which correspond to the most challenging and critical aspects (Komatsu and Ahsan, 2009).

Subproteomes

Since the preparation protocols do not allow evaluation of the complete proteome and peptidome, the protocol definition should take into consideration protein subgroups and source material of interest, whether it is an organelle, a cell or a tissue, in order to reduce sample complexity, enrich and increase the possibility of identifying LAP of interest involved in different cellular mechanisms and different locations. Hence, the information obtained for each subproteome of the same source material contributes to achieve greater coverage of the proteome being studied. This strategy has attracted the attention of several researchers (Cánovas *et al.*, 2004; Zhang *et al.*, 2004; Natarajan *et al.*, 2005, 2009; Oehrle *et al.*, 2008; Jorrin-Novo *et al.*, 2009; Krishnan and Natarajan, 2009; Agrawal *et al.*, 2010; Matros *et al.*, 2010; Kota and Goshe, 2011).

Plant cell wall proteins

For the study of plant proteomes, the extracts may be successfully fractionated into soluble proteins and cell wall (CW) proteins (Figure 1). CW proteins represent a subproteome of great importance, since many of these molecules are involved in the maintenance of the cellular structure and in processes of plant defense, for instance in responses to abiotic and biotic stresses (Jamet *et al.*, 2006; Kong *et al.*, 2010), or as a constitutive barrier against pathogenic microorganisms. Studies in our laboratory evidenced high constitutive antimicrobial activity against two plant-pathogenic bacteria by the peptide fraction obtained from CW extracts from leaves of bell pepper (Figure 2, Teixeira *et al.*, 2006) and 60-day-old eggplant (Almeida *et al.*, 2008). Also for eggplants, the highest inhibition levels were obtained with CW extracts from 5-cm-tall plants, while soluble extracts promoted the highest inhibition rate when fully expanded leaves were analyzed (Table 1). These results suggested that young plants exhibit an innate defense mechanism, very likely to minimize plant microbial invasion, whereas expanded leaves produce soluble defense molecules (Almeida *et al.*, 2007).

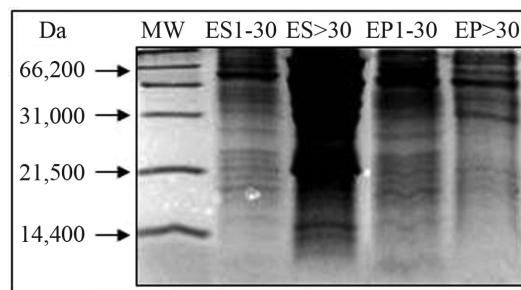


Figure 1 - Denaturing electrophoresis (16.5%) of soluble extracts (SE) and cell wall extracts (CWE) of leaves of soybean genotypes 'PI 561356' (Embrapa CNPSo) subjected to injury from seven wounds on the leaves. SE and CWE were fractionated by ultrafiltration and molecules contained between 1 and 10 kDa (1-10) and greater than 30 kDa (> 30). MM: 5 μ L of the Wide-Range Molecular Weight Markers (Bio-Rad, USA). The peptide bands were stained with Coomassie Brilliant Blue R250. (Courtesy of José Fabiano Sena-Netto, from the final paper for his undergraduate degree in Biochemistry in 2010 at the Federal University of Viçosa, Brazil).

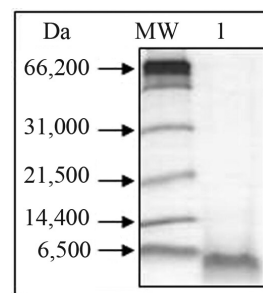


Figure 2 - Three-phase SDS-Tricine-PAGE of Cell Wall Extracts (CWE) from fully expanded leaves of Samara bell pepper variety, after ammonium sulfate precipitation followed by heating fractionation and dialysis (line 1). MM was 5 μ L of the Wide-Range Molecular Weight Markers (Bio-Rad, USA). The protein bands were stained with Coomassie Brilliant Blue G250.

Table 1 - Antimicrobial activity of cationic (CF1, CF2) and anionic (AF) fractions obtained from soluble (SE) or cell-wall (CWE) extracts of young plants or fully-expanded leaves of eggplant (Almeida *et al.*, 2007, with authorization from the authors).

Peptide-enriched fractions	Bacterial growth inhibition (%)*			
	<i>Ralstonia solanacearum</i>			
	5 cm whole plants (1.25 g)**	15 cm whole plants (1.25 g)**	Fully-expanded leaves	
			(0.75 g)**	(1.5 g)**
SE-CF1	-43.3 ± 0.082***	92.2 ± 0.003	90.8 ± 0.004	85.2 ± 0.000
SE-CF2	0.75 ± 0.012	86.2 ± 0.001	100 ± 0.003	100 ± 0.001
SE-AF	23.9 ± 0.005	14.2 ± 0.001	92.0 ± 0.000	100 ± 0.005
CWE-CF1	8.60 ± 0.000	3.70 ± 0.014	1.80 ± 0.001	45.8 ± 0.029
CWE-CF2	91.0 ± 0.000	80.0 ± 0.004	29.6 ± 0.019	88.0 ± 0.006
CWE-AF	14.2 ± 0.005	0.75 ± 0.007	69.0 ± 0.006	73.4 ± 0.018

*The percentage values refer to the average ± SD, for two replications; ** Extract concentration refers to 1.25 g of fresh young plant material, or 0.75 g and 1.5 g of fresh fully-expanded leaf, with cultivation for 14 h; *** The negative signal indicates the occurrence of larger growth than in the culture-control.

Plant CWs are highly dynamic and contain chemically active compounds secreted by the cells, which are essentially polysaccharides and proteins, the latter comprising approximately 10% of the CW mass (Jamet *et al.*, 2008). These proteins, which are difficult to isolate from these complex matrixes, require specific extraction methods (Teixeira *et al.*, 2006; Zhu *et al.*, 2006; Almeida *et al.*, 2007, 2008; Kong *et al.*, 2010), due to their low abundance. Differential extraction enriches the extract, allows access to the CW LAP, and facilitates the comparison of the expression profiles under different stress conditions (Watson and Summer, 2006; Negri *et al.*, 2008).

Protein extraction from CWs can be achieved by methods that may involve or not cell disruption. For each procedure there are advantages and disadvantages, especially concerning contamination and experimental procedures. Methods commonly used involve calcium chloride and lithium chloride solutions (Feiz *et al.*, 2006; Kong *et al.*, 2010). Calcium chloride does not disrupt the cell wall but is capable of releasing ionic molecules externally anchored to the CW, whereas lithium chloride solution is used to extract proteins intrinsic to the CW, such as glycoproteins (Feiz *et al.*, 2006). When comparatively evaluating the CW subproteomes of soybean leaves submitted or not to biotic stress using reversed-phase high performance liquid chromatography we obtained a group of differential protein peaks in the comparative chromatograms, demonstrating that sample complexity was reduced. This procedure allows the identification of the isolated proteins.

Proteins with different ranges of molecular masses

When the aim is to identify proteins of a specific group of molecular mass present in a complex sample, methodologies based on ultrafiltration have an advantage, since they are capable of fractionating, concentrating and exchanging solutions, and thus recovering native samples with molecular masses of interest. Ultrafiltration is a low-

cost process requiring simple equipment only, such as filtering membranes of different cut-off values, which allow the fractionation of different protein groups. Soluble (SE) and CW extracts (CWE) of soybean leaves submitted (Figure 1) or not to stress were successfully fractionated in our laboratory through ultrafiltration. Fractions for peptidomics and proteomics, with proteins smaller than 30 kDa and greater than 30 kDa, respectively, were recovered. After separating the fractions by reversed-phase chromatography (RPC) and recovering the proteins and peptides for mass spectrometry analysis, differentially expressed molecules were successfully identified.

Molecular exclusion chromatography (MEC) is also an alternative used by our group for this purpose. This method favors isolating a specific protein group according to the selected resin. In general, MEC presents low accuracy in the separation of small proteins and peptides, yet it is a fast procedure that generates clean samples with low salt concentration, which is important for later stages of purification, for instance during RPC. MEC has been successfully used in our proteomic analyses using multi-dimensional liquid chromatography. A constitutive antimicrobial peptide was identified in CW extracts from tomato plant after MEC following RPC while an induced defense peptide was detected in plants after abiotic stress (submitted for publication).

Proteins from different cellular compartments

Ultra-centrifugation enables the separation of cellular organelles according to their sizes and densities. Zonal centrifugation or centrifugation by differential velocity may be used for fractionation, the latter being a technique which generates greater sedimentation strength, enabling the separation of organelles of similar characteristics (Komatsu, 2006), such as nuclei, mitochondria or plastids,

while maintaining their integrity (Wijk, 2004; Kosová *et al.*, 2011).

Contaminant removal and use of additives

A limiting factor for proteomic studies is the presence of protein and non-protein contaminants in the samples. Contaminants may interfere with the detection, isolation or identification of proteins of interest, acting by aggregation with the proteins, by affecting the signal/noise relationship of the detection equipment, by promoting protein degradation or reducing the activity of enzymes, among other factors. Under native conditions, the removal of these contaminants may be achieved by dialysis, salt fractionation, or with organic cold solvents, through ultrafiltration or preparative chromatography. Under denaturing conditions, steps including selective heating or addition of concentrated acids may favor sample bleaching and the enrichment of the protein group of interest. Also, anionic detergents like sodium dodecyl sulfate (SDS) or nonionic detergents such as Triton X-100 and Nonidet P-40 can be added in different concentrations. The concentration to be used must be empirically defined.

Fractionation by ultrafiltration allows the simultaneous desalting and concentration of the samples in the native form (Kong *et al.*, 2010). However, during the ultrafiltration procedure certain caution is required so as to maintain the sample in its native form, as well as to avoid protein agglomeration. Some of the factors to be controlled are hydrophobicity, pH and ionic strength of the buffer or solution being used. The choice of salts and pH of the buffer solution for ultrafiltration depends on the subproteome under study. For instance, as pH values around 5.5 are close to the pH of the cell wall, the fractionation of CW proteins should be done at pH values close to 5.5 to avoid precipitation of these proteins (Watson and Summer, 2006). On the other hand, many storage proteins, which are usually contaminants, precipitate between pH 4.5 and 4.8 (Speroni *et al.*, 2010), and extraction at these pH values produces less complex extracts. For ultrafiltration of membrane proteins and partially hydrophobic proteins, detergents should be used, as these become soluble in the presence of amphipathic compounds, such as 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), SDS and the nonionic surfactant Triton X-100, among others.

Pressure is considered an important factor which frequently is not controlled during ultrafiltration. There are not sufficient studies reporting protein agglomeration, solubility and protein conformation under pressure (Chalikian and MacGregor, 2009; Speroni *et al.*, 2010). Iwabuchi and Yamauchi (1987), fractionating glycinin (360 kDa) and β -conglycinin (180 kDa), found low molecular mass proteins in the protein fraction containing these storage proteins, suggesting the occurrence of agglomeration of low

molecular mass proteins together with high molecular mass proteins.

In our experience, when a low salt concentration buffer (5 mM Tris buffer, pH 7.0) was used after the ultrafiltration procedure, the occurrence of agglomeration in extracts from soybean leaves increased. Agglomeration was also increased under cold temperatures, and an insoluble precipitate was observed in fractions containing low and high molecular masses. Replacing this buffer with 20 mM Tris-HCl, pH 7.0, containing 20 mM ammonium acetate, agglomeration was no longer observed, suggesting that agglomeration was an ionic behavior. In an attempt to disaggregate the precipitate formed in the high molecular fraction, different procedures were used. Disaggregation was not achieved in the presence of 0.5% (v/v) Triton X-100 or by dilution (up to 10 times) with 100 mM Tris-HCl, pH 7.0, but the sequential addition of 0.4, 1.0 and 2.0% (w/v) SDS partially reduced the agglomeration, thus corroborating the hypothesis of ionic behavior.

Denaturing one-dimensional electrophoresis (16.5% T) done with extracts from soybean leaves enriched for low molecular mass proteins (up to 30 kDa) yielded protein bands higher than 30 kDa, suggesting the occurrence of agglomeration after ultrafiltration. This was confirmed when a 60-kDa band was eluted from the gel and again separated under the same conditions, this time producing 6-kDa and 30-kDa bands. Similarly, the elution and new separation of a 6-kDa band has also produced the same bands, confirming that heating the sample for 10 min in the sample buffer (Teixeira *et al.*, 2006) was unable to solubilize proteins in the sample.

Fractionation with ammonium sulfate has been used by our group since 2000 for bleaching plant samples, to remove contaminants such as carbohydrates, phenolic compounds and pigments, among others. Plant extracts obtained in the absence of this compound presented high noise/signal relationship in the spectrometry analyses and protein agglomeration was observed. Recently, Park *et al.* (2008) stressed the importance of this procedure to remove interfering molecules. Nevertheless, caution should be taken, as the presence of salt may interfere with electrophoretic properties and the MS procedure.

Specific stages of the proteomic analysis require samples with specific characteristics. 2-DE, for instance, is highly affected by salts, as these may accumulate in the tips of the strips during the first-dimension separation. They can cause heat generation and the formation of zones with different conductivities and degrees of hydration. During MS analysis, salts may increase noise that interferes with the ionization of the molecules. SDS interferes with the RFC procedure by causing the degradation of the resins. Samples with high protein concentration may precipitate and those with low protein concentration may prevent the detection of LAP. Hence, samples should be prepared in buffers with low ionic strength and in the presence of protease

inhibitors with different specificities (unpublished date). The use of non-interfering additives is an empirical decision, which should be carefully evaluated because these compounds may have to be removed in a subsequent purification step.

Refined methods for isolation

Separation by two-dimensional electrophoresis

Two-dimensional electrophoresis (2-DE) is the most commonly used separation technique in proteomics for comparative and global protein analyses under differential conditions. A high number of proteins can be identified in a single gel. Considering that in general, plant tissues do not present a high protein concentration, and that the presence of proteases and interfering molecules can drastically affect proteomic analyses, an efficient protein extraction protocol must be used. This protocol should be able to eliminate secondary metabolites, remove additives which are not compatible with the different stages of purification, and enrich the LAP. Extraction methods using phenol in conjunction with ammonium acetate/methanol precipitation have proved to be highly efficient (Isaacson *et al.*, 2006). Using this procedure Xu *et al.* (2006) were able to obtain a large number of intense and well-resolved spots from soybean leaves in 2-D gels. Similar results were obtained in our laboratory. Methods based on precipitation with trichloroacetic acid (TCA) have also been used in the literature (Chen and Harmon, 2006) and in our laboratory. Using a method based on an initial TCA precipitation step followed by extraction by the dense phenol/SDS method and a final precipitation with ammonium acetate/methanol, Wang *et al.* (2003) obtained a high number of soluble and membrane proteins from soybean leaves. To optimize this method, we introduced some modifications, such as the use of protease inhibitors (1 mM phenylmethylsulfonyl fluoride), reducing agents (different concentrations of β -mercaptoethanol and dithiothreitol) and 1% polyvinylpyrrolidone, which eliminate or prevent the action of proteases and compounds that interfere with the subsequent steps in protein separation and identification.

Liquid chromatography (LC) can be used as an alternative to 2-DE when the proteins of interest present extremes of molecular mass and pI, when they are highly hydrophobic or LAP (Natarajan *et al.*, 2009; Gilmore and Washburn, 2010; Kolodziejek and van der Hoorn, 2010).

Separation by chromatographic methods

Chromatographic methods using a combination of different separation principles (multi-dimensional) in on-line or off-line processes, or the selective separation by affinity have been used and improved to resolve the complexity of the extracts and/or to enrich them. Commercially available on-line chromatographic systems generally use a strong ionic exchange column (IEC) followed by a reverse phase. Molecular exclusion chromatography (MEC) fol-

lowed by reverse phase chromatography (RPC) is also used (Issaq, 2001), though less often. Our group has worked with off-line chromatography with success, especially associating MEC and RPC for prospection of proteins with different ranges of molecular masses, when two antimicrobial peptides from tomato leaves were purified under native conditions and subsequently identified (data submitted for publication). These procedures are orthogonal, creating a large number of fractions to be analyzed, making proteomic analyses by MDLC difficult and lengthy. However, it allows a greater proteome and peptidome prospection, with little restriction for very large or very small and electrically charged membrane proteins. And in particular, it allows the isolation of native purified samples for evaluation of the proteome functionality.

Although these methods are very efficient, complementary techniques are required for obtaining further information regarding the proteome. Thus, the development and application of technologies and methods for separation by affinity and enrichment have become a high priority for defining separomic protocols to be used for different organisms.

A great variety of affinity and immuno-affinity columns are already available and allow the separation of complex biological extracts into different protein classes, or even to fractionate LAP. These commercially available columns are important for the identification of protein-protein interaction networks (Azarkan *et al.*, 2007) and to remove unwanted HAP proteins, peptides and nucleotides, metals, etc. (Keshishian *et al.*, 2007; Cellar *et al.*, 2008; Fang and Zang, 2008; Huang and Fang, 2008; Miernyk and Hajduch, 2011).

Identifying Soybean Proteins and Peptides By Proteomics Tools

For comparative proteomics and peptidomics, 2-DE is the proteomic platform giving highest-throughput (Brandão *et al.*, 2010). More sophisticated procedures, such as 2-D DIGE and isotopic labeling can be used. In most soybean proteomic studies using 2-D procedures, the proteins undergo electrophoresis, are trypsinized in gel and are then analyzed by MALDI-MS/MS. The PMF technique (Peptide Mass Fingerprint) following MS correlates the peptide masses after proteolysis of the unknown protein with the peak list of theoretical proteins deposited in public databases (Chamrad *et al.*, 2004; Elias *et al.*, 2005), which are assessed by search software such as MASCOT (Thiede *et al.*, 2005). Similarly, PFF (Peptide Fragment Fingerprint) correlates mass spectra of fragments of these peptides (Chamrad *et al.*, 2004; Elias *et al.*, 2005; Nielsen *et al.*, 2005). Fragmentation spectra can also be analyzed by “de novo” peptide sequencing using MS, which gives the amino acid sequences (Roepstorff and Fohlman, 1984). The development of a plethora of bioinformatics resources has

allowed the identification and characterization of proteins with accuracy, thanks to genomic and protein sequence information deposited in public databases (Schmutz *et al.*, 2010).

Conclusions

Proteomics is an invaluable tool for functional genomic analysis of plants. Techniques with different sensitivities and accuracies should be used in the analysis of the respective proteome. Highly sensitive and accurate identification equipment is available; however, better protein and peptide extraction procedures and methods for enrichment and isolation of low-abundance proteins with transient expression and of rare peptides are needed. Some alternatives to overcome or minimize these limitations are: (1) the decrease of extract complexity to allow evaluation of subproteomes, (2) use of methodologies and/or additives to remove highly abundant proteins known as contaminants, (3) use of chromatographic techniques which allow the enrichment of specific protein fractions, and (4) isolation of proteins in their native form to aid their identification and validation of metabolic pathways in which they are involved. The use of basic biochemistry tools which separate protein and peptides based on their physico-chemical characteristics can further the enrichment process. Knowledge about the chemical and structural characteristics of non-protein contaminants is helpful in the choice of methods and tools for their removal or inactivation. This information will also aid in defining methods to prevent solubilization of these contaminants and their consequent contact with the proteins of interest.

Acknowledgments

The authors wish to express their gratitude to the Brazilian agencies CNPq, FAPEMIG, FINEP and CAPES for financial support and fellowships, and to Embrapa Soja and NuBioMol/UFV for partnership and technical support.

References

- Almeida HO, Mattos EC, Barbosa MO, Teixeira FR, Magalhaes RDM, Romeiro RS, Fontes EPB and Baracat-Pereira MC (2007) Peptide fraction inhibiting plant pathogen growth predominated in cell wall extracts from young plants or in soluble cell fraction from expanded leaves from eggplants. *J Phytopathol* 155:735-737.
- Almeida HO, Teixeira FR, Romeiro RS, Silva DJH, Pereira PRG, Fontes EPB and Baracat-Pereira MC (2008) Atividade antimicrobiana de extratos peptídicos de folhas de berinjela na inibição do crescimento de *Ralstonia solanacearum* e *Clavibacter michiganensis* subsp. *michiganensis*. *Summa Phytopathol* 34:62-64 (Abstract in English).
- Azarkan M, Huet J, Baeyens-Volant D, Looze Y and Vandebussche G (2007) Affinity chromatography: A useful tool in proteomics studies. *J Chromatogr B Analyt Technol Biomed Life Sci* 849:81-90.
- Bindschedler LV and Cramer R (2011) Quantitative plant proteomics. *Proteomics* 11:756-775.
- Brandão AR, Barbosa HS and Arruda MA (2010) Image analysis of two-dimensional gel electrophoresis for comparative proteomics of transgenic and non-transgenic soybean seeds. *J Proteomics* 73:1433-1440.
- Cánovas FM, Dumas-Gaudot E, Recorbet G, Jorin J, Mock H-P and Rossignol M (2004) Plant proteome analysis. *Proteomics* 4:285-298.
- Carpentier SC, Witters E, Laukens K, Deckers P, Swennen R and Panis B (2005) Preparation of protein extracts from recalcitrant plant tissues: An evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics* 5:2497-2507.
- Cellar NA, Kuppanan K, Langhorst ML, Ni W, Xu P and Young SA (2008) Cross species applicability of abundant protein depletion columns for ribulose-1,5-bisphosphate carboxylase/oxygenase. *J Chromatogr B Analyt Technol Biomed Life Sci* 861:29-39.
- Chalikian TV and MacGregor Jr RB (2009) Origins of pressure-induced protein transitions. *J Mol Biol* 394:834-842.
- Chamrad DC, Korting G, Stuhler K, Meyer HE, Klose J and Bluggel M (2004) Evaluation of algorithms for protein identification from sequence databases using mass spectrometry data. *Proteomics* 4:619-628.
- Chen S and Harmon AC (2006) Advances in plant proteomics. *Proteomics* 6:5504-5516.
- Cho JH, Hwang H, Cho M-H, Kwon Y-K, Jeon J-S, Bhoo SH and Hahn T-R (2008) The effect of DTT in protein preparations for proteomic analysis: Removal of a highly abundant plant enzyme, ribulose bisphosphate carboxylase/oxygenase. *J Plant Biol* 51:297-301.
- Cho WCS (2007) Review – Proteomics technologies and challenges. *Genomics Proteomics Bioinform* 5:77-85.
- Corthals GL, Wasinger VC, Hochstrasser DF and Sanchez JC (2000) The dynamic range of protein expression: A challenge for proteomic research. *Electrophoresis* 21:1104-1115.
- Elias JE, Haas W, Faherty BK and Gygi SP (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Meth* 2:667-675.
- Fang X and Zhang W (2008) Affinity separation and enrichment methods in proteomic analysis. *J Proteomics* 71:284-303.
- Feiz L, Irshad M, Pont-Lezica RF, Canut H and Jamet E (2006) Evaluation of cell wall preparations for proteomics: A new procedure for purifying cell walls from *Arabidopsis* hypocotyls. *Plant Meth* 2:e10.
- Gilmore JM and Washburn MP (2010) Advances in shotgun proteomics and the analysis of membrane proteomes. *J Proteomics* 73:2078-2091.
- Huang L, Harvie G, Feitelson JS, Gramatikoff K, Herold DA and Allen DL (2005) Immunoaffinity separation of plasma proteins by IgY microbeads: Meeting the needs of proteomic sample preparation and analysis. *Proteomics* 5:3314-3328.
- Huang L and Fang X (2008) Immunoaffinity fractionation of plasma proteins by chicken IgY antibodies. *Meth Mol Biol* 425:41-51.
- Isaacson T, Damasceno CMB, Saravanan RS, He Y, Catalá C, Saladié M and Rose JKC (2006) Sample extraction techniques for enhanced proteomic analysis of plant tissues. *Nat Prot* 1:769-774.

- Issaq HJ (2001) The role of separation science in proteomics research. *Electrophoresis* 22:3629-3638.
- Iwabuchi S and Yamauchi F (1987) Electrophoretic analysis of whey proteins present in soybean globulins fraction. *J Agric Food Chem* 35:205-209.
- Jamet E, Canut H, Boudart G and Pont-Lezica RF (2006) Cell wall proteins: A new insight through proteomics. *Trends Plant Sci* 11:33-39.
- Jamet E, Albenne C, Boudart G, Irshad M, Canut H and Pont-Lezica R (2008) Recent advances in plant cell wall proteomics. *Proteomics* 8:893-908.
- Jorrín-Novo JV, Maldonado AM, Echevarría-Zomeño S, Villedor L, Castillejo MA, Curto M, Valero J, Sghaier B, Donoso G and Redondo I (2009) Plant proteomics update (2007-2008): Second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge. *J Proteomics* 72:285-314.
- Jurgens M and Schrader M (2002) Peptidomic approaches in proteomic research. *Curr Opin Mol Ther* 4:236-241.
- Keshishian H, Addona T, Burgess M, Kuhn E and Carr SA (2007) Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics* 6:2212-2229.
- Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY, Kim DS, Lee Y S, Park D, Ma J, *et al.* (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proc Natl Acad Sci USA* 107:22032-22037.
- Kolodziejek I and van der Hoorn RAL (2010) Mining the active proteome in plant science and biotechnology. *Curr Opin Biotechnol* 21:225-233.
- Komatsu S (2006) Extraction of nuclear proteins. In: Thiellement H, Zivy M, Damerval C and Méchin V (eds) *Methods in Molecular Biology*, vol. 335: *Plant Proteomics: Methods and Protocols*. Humana Press Inc, Totowa, pp 73.
- Komatsu S and Ahsan N (2009) Soybean proteomics and its application to functional analysis. *J Proteomics* 72:325-336.
- Kong F-J, Oyanagi A and Komatsu S (2010) Cell wall proteome of wheat roots under flooding stress using gel-based and LC MS/MS-based proteomics approaches. *Biochim Biophys Acta* 1804:124-136.
- Kosová K, Vítámvás P, Prásil IT and Renaut J (2011) Plant proteome changes under abiotic stress – contribution of proteomics studies to understanding plant stress response. *J Proteomics* 74:1301-1322.
- Kota U and Goshe MB (2011) Advances in qualitative and quantitative plant membrane proteomics. *Phytochemistry* 72:1040-1060.
- Krishnan HB and Natarajan SS (2009) A rapid method for depletion of Rubisco from soybean (*Glycine max*) leaf for proteomic analysis of lower abundance proteins. *Phytochemistry* 70:1958-1964.
- Lescuyer P, Hochstrasser D and Rabilloud T (2007) How shall we use the proteomics toolbox for biomarker discovery? *J Proteome Res* 6:3371-3376.
- Matros A, Kaspar S, Witzel K and Mock HP (2010) Recent progress in liquid chromatography-based separation and label-free quantitative plant proteomics. *Phytochemistry* 72:963-974.
- Miernyk JA and Hajduch M (2011) Seed proteomics. *J Proteomics* 74:389-400.
- Natarajan S, Xu C, Caperna TJ and Garrett WM (2005) Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. *Anal Biochem* 342:214-220.
- Natarajan SS, Krishnan HB, Lakshman S and Garrett WM (2009) An efficient extraction method to enhance analysis of low abundant proteins from soybean seed. *Anal Biochem* 394:259-268.
- Negri AS, Prinsi B, Scienza A, Morgutti S, Cocucci M and Spen L (2008) Analysis of grape berry cell wall proteome: A comparative evaluation of extraction methods. *J Plant Physiol* 165:1379-1389.
- Nielsen ML, Savitski MM and Zubarev RA (2005) Improving protein identification using complementary fragmentation techniques in Fourier transform mass spectrometry. *Mol Cell Proteomics* 4:835-845.
- Nouri MZ and Komatsu S (2010) Comparative analysis of soybean plasma membrane proteins under osmotic stress using gel-based and LC MS/MS-based proteomics approaches. *Proteomics* 10:1930-1945.
- Oehrle NW, Sarma AD, Waters JK and Emerich DW (2008) Proteomic analysis of soybean nodule cytosol. *Phytochemistry* 69:2426-2438.
- Park J-W, Lee S-G, Song JY, Joo JS, Chung M-J, Kim S-C, Youn H-S, Kang H-L, Baik S-C, Lee W-K, *et al.* (2008) Proteomic analysis of *Helicobacter pylori* cellular proteins fractionated by ammonium sulfate precipitation. *Electrophoresis* 29:2891-2903.
- Roepstorff P and Fohlman J (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 11:601.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, *et al.* (2010) Genome sequence of the paleopolyploid soybean (*Glycine max* (L.) Merr.). *Nature* 463:178-183.
- Speroni F, Añón MC and Lamballerie M (2010) Effects of calcium and high pressure on soybean proteins: A calorimetric study. *Food Res Int* 43:1347-1355.
- Sudaric A, Vrataric M, Drinic SM and Matosa M (2010) "Biotechnology" in soybean breeding. *Genetika-Belgrade* 42:91-102.
- Teixeira FR, Lima MCO, Almeida HO, Romeiro RS, Silva DJH, Pereira PRG, Fontes EPB and Baracat-Pereira MC (2006) Bioprospection of cationic and anionic antimicrobial peptides from bell pepper leaves for inhibition of *Ralstonia solanacearum* and *Clavibacter michiganensis* ssp. *michiganensis* growth. *J Phytopathol* 154:418-421.
- Thiede B, Höhenwarter W, Krah A, Mattow J, Schmid M, Schmidt F and Jungblut PR (2005) Peptide mass fingerprinting. *Methods* 35:237-247.
- Vertommen A, Moller ALB, Cordewenerd JHG, Swennen R, Panis B, Finnie C, America AHP and Carpentiera SC (2011) A workflow for peptide-based proteomics in a poorly sequenced plant: A case study on the plasma membrane proteome of banana. *J Proteomics* 74:1218-1229.
- Wang W, Scali M, Vignani R, Spadafora A, Sensi E, Mazzuca S and Cresti M (2003) Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing

- high levels of interfering compounds. *Electrophoresis* 24:2369-2375.
- Watson BS and Summer LW (2006) Isolation of cell wall proteins from *Medicago sativa* stems. In: Thiellement H, Zivy M, Damerval C and Méchin V (eds) *Methods in Molecular Biology*, vol. 335: *Plant Proteomics: Methods and Protocols*. Humana Press Inc, Totowa, pp 79.
- Wijk KJ (2004) Plastid proteomics. *Plant Physiol Biochem* 42:963-977.
- Xu C, Garrett WM, Sullivan J, Caperna TJ and Natarajan S (2006) Separation and identification of soybean leaf proteins by two-dimensional gel electrophoresis and mass spectrometry. *Phytochemistry* 67:2431-2440.
- Yamaguchi M and Sharp RE (2010) Complexity and coordination of root growth at low water potentials: Recent advances from transcriptomic and proteomic analyses. *Plant Cell Environ* 33:590-603.
- Zhang H, Yan W and Aebersold R (2004) Chemical probes and tandem mass spectrometry: A strategy for the quantitative analysis of proteomes and subproteomes. *Curr Opin Chem Biol* 8:66-75.
- Zhu J, Chen S, Alvarez S, Asirvatham VS, Schachtman DP, Wu Y and Sharp R (2006) Cell wall proteome in the maize primary root elongation zone. I. Extraction and identification of water-soluble and lightly ionically bound proteins. *Plant Physiol* 140:311-325.

Internet Resources

Phytozome ver. 7.0: *Glycine max*. Center for Integrative Genomics (CIG), Joint Genome Institute (JGI), University of California Regents, <http://www.phytozome.net/soybean> (January 24, 2011).

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.