PLANT PROTEOMICS: CHALLENGES AND CONTRIBUTIONS FOR GENETIC IMPROVEMENT

Angela Mehta[1]

In recent years, functional genomic strategies, including proteomics and transcriptomics, have been extensively used in an effort to define gene and protein function. Currently, large amounts of data are available due to the completion of several genome sequencing projects. Proteomics, or the analysis of the protein complement of the genome, provides experimental continuity between genome sequence information and the protein profile in a specific tissue, cell, or cellular compartment during standard growth or under different treatment conditions. While the genome defines potential contributions to cellular function, the expressed proteome represents actual contributions. Moreover, by using proteomic approaches, differences in the abundance of proteins present at the time of sampling can be distinguished and different forms of the same protein can be resolved (MEHTA et al., 2008a).

It is well known that the levels of mRNA do not necessarily predict the levels of the corresponding proteins in the cell (JONES et al., 2004). Differing stability of mRNAs and different efficiencies in translation can affect the generation of new proteins, which makes proteomic investigation even more important. Post-translational modifications such as the removal of signal peptides, phosphorylation, glycosilation, ubiquitination, among others are also important processes for protein function which are not considered when using the genomic approach (JENSEN, 2006). Therefore, proteomics is playing an increasingly important role in addressing these issues and has become a necessary and complementary approach in the post-genomic era. Furthermore, by analyzing the proteins being expressed during a specific condition, information regarding the genes/proteins that are co-regulated and act together in response to a given stress can be identified.

Proteomics has dramatically evolved in pursuit of large-scale function assignment of candidate proteins. The application of proteomic approaches for global expression analysis and protein identification has been highly efficient in different fields of investigation. These analyses have been performed by exploring the high resolution of two-dimensional electrophoresis (2-DE) coupled to mass spectrometry. These data, when complemented by *de novo* sequencing, allows

[1] Embrapa Recursos Genéticos e Biotecnologia, PqEB AV W/5 Norte Final, CEP 70770-900, Brasília-DF, Brazil

the unequivocal identification of proteins involved in different biological functions. Some authors have suggested that 2-DE is an ancient and surpassed technique, although this procedure has been frequently utilized in diverse research areas, including plant proteomics. In fact, promising alternative technologies such as multidimensional protein separation, protein arrays and others have emerged recently (MEHTA et al., 2008b). However, 2-DE is currently the most feasible technique that can be routinely applied for quantitative expression profiling of large sets of complex protein mixtures (GORG et al., 2004). Although 2-DE technology is not properly cheap, the equipments are much more accessible than mass spectrometers, for example. This particularity makes proteomic studies possible for several research groups located in developing countries. It is true that protein identification is an important step in proteomic analysis and mass spectrometry-based strategies have contributed enormously in this aspect. Thus, the recent enthusiasm in proteomic studies is a result of the union of techniques based on 2-DE and those focused on mass spectrometry, and this union has been responsible for the increase in the functional assignment of proteins and genes in various organisms, including plants (MEHTA et al., 2008b).

The 2-DE technology began to be explored in the 1970s (O'FARREL, 1975) and several modifications have been made to this classical technique. Immobilized pH gradients and gels, as well as power supplies and gel supports are now commercially available and have clearly improved reproducibility. Recent developments and technical advances in mass spectrometry were also achieved and these improvements have lead to a high throughput analysis of protein expression in different organisms. However, in plants, proteomic research has not yet reached a sufficient level of complexity, when compared to the advances obtained in prokaryotes, yeasts and mammalians (AGRAWAL & RAKWAL, 2006), for example. Although several proteomic studies have been performed in plants in order to search for gene products of agronomic relevance, the data available is still limited, especially considering the different tissues. Root proteomic studies, for example, are highly insipient when compared to other tissues. Therefore plant proteomics can be considered an interesting field of investigation, which needs to advance considerably in order to keep pace with the research in other organisms.

The proteomic approach has been used as a fundamental method to understand and identify the functions of proteins expressed in a given condition, however, some important limitations need to be cosidered and may account for the slower advance of proteomics when compared to genomics. Recently, our group presented a review discussing proteomics of plant-pathogen interactions (MEHTA et al., 2008a). We showed that the data available revealed several proteins commonly expressed in diverse pathosystems. In the case of the pathogens,

several proteins involved in pathogenicity comprised secretion proteins, which were observed in bacteria, fungi and nematodes and were mainly identified by secretomic studies. These proteins included proteases, cellulases and pectate lyases, which are important cell wall degrading enzymes (CWDE), crucial for host plant colonization. Besides these well-known enzymes, other proteins associated to protection against oxidative stress response by the plant upon infection were also reported in the different pathogens. A similar scenario was observed regarding defence-related proteins in plants. The most reported defence-related proteins were PR proteins such as thaumatins, glucanases, peroxidases and chitinases observed in several pathosystems (MEHTA et al., 2008a).

Although several proteins expressed during plant-pathogen interactions were highlighted, most of them were well known. The results obtained from most proteomic analyses are of extreme importance for validation of the expression of the genes identified by genomic or transcriptomic studies. However, a low amount of novel information has been obtained and could be explained by the fact that key proteins are expressed in low abundance, and are therefore not detected by current proteomic tools (MEHTA et al., 2008a). Indeed only the most abundant proteins are usually detected in 2D gels and successfully identified by mass spectrometry.

Another major problem faced in proteomic analyses is protein identification by mass spectrometry. Unequivocal identification is usually obtained only when the genome sequence or a high amount of sequence data is available in the public databases. When analyzing poorly studied organisms, identification must be performed by *de novo* sequencing, which requires more sophisticated equipments and a laborius data interpretation. Therefore, a gap appears to exist in the bioinformatics pipeline for proteomics of organisms with incomplete sequenced genomes. These technical limitations in proteomic studies need to be overcome in order to advance our knowledge on protein expression (MEHTA et al., 2008a).

In spite of these limitations, we believe that proteomics can privide a considerable amount of information in protein expression and can help understand several biological processes in plants. Regarding our research in the proteomic field, we have investigated protein expression under drought conditions in upland rice. The comprehension of drought responses in upland rice is important for designing breeding strategies to develop varieties more tolerant to water constraints. Although numerous genes and proteins, which potentially contribute to drought tolerance in rice, have been previously reported (YANG et al., 2004; FU et al., 2007; GORANTLA et al., 2007; WU et al., 2006), most of these studies focused on lowland rice genotypes. Moreover, most ESTs from drought stressed plants available were obtained from

libraries constructed using seedlings (REDDY et al., 2002). We have analyzed drought-stressed plants in the reproductive stage and used root tissue of plants grown under defined drought conditions.

Stressed plants from both genotypes were submitted to drought stress after anthesis and the roots were collected after twenty-one days of stress. Triplicates of the gels from each genotype were compared and revealed a total of 463 proteins in the Prata Ligeiro (tolerant genotype) profile and 522 in IRAT20 (susceptible genotype). The two obtained synthetic gels were overlapped and this procedure allowed the identification of 307 overlapped spots, 156 proteins exclusive to the tolerant genotype and 215 proteins exclusive to the susceptible genotype (RABELLO et al., 2008).

A total of 50 intense proteins observed in the tolerant genotype profile after Coomassie blue staining was excised from the gel, digested and analyzed by mass spectrometry. By using the Mascot program, 22 proteins could be identified with a significant score, including 16 up- and 4 down-regulated, 1 new and 1 equally expressed in both genotypes (RABELLO et al., 2008). The other proteins were in insufficient amounts for the identification analysis or did not return reliable matches when using the Mascot program. This probably occurs due to a low protein quantity and/or low ionization capacity of molecular components present in the samples analyzed. Among the identified proteins were several hypothetical proteins and proteins involved in oxidative stress protection such as superoxide dismutase [Cu-Zn], L- ascorbate peroxidase and ascorbate peroxidase (RABELLO et al., 2008). The expression of genes associated with cell protection against oxidative damage is considered important to cope with water deficit in upland rice and proteins related to this function presented a higher expression in the tolerant genotype (RABELLO et al., 2008). We proposed that the upland susceptible genotype responds to drought in a similar way as lowland rice, which is naturally more susceptible to water stress (RABELLO et al., 2008).

Another example of our reasearch involves a plant-pathogen interaction. We have investigated the protein expression of cowpea beans (*Vigna unguiculata*) during nematode infection. Cowpea plants were inoculated with approximately 5000 J2 of *M. incognita* and the roots were collected at 3, 6 and 10 days after inoculation. Protein extraction was performed with phenol according to de Mot & Vanderleyden (1989). The 2-DE analysis revealed approximately 300 proteins per gel and 26 differentially expressed proteins. The resistant genotype presented 13 up- and 10 down-regulated and 3 exclusive protein spots. The differentially expressed proteins were excised form the gel and analyzed by mass spectrometry. A total of 12 proteins were sucessfully identified and some of them may be involved in the resistance process.

Taken together, we have mentioned some examples of the application of proteomic studies in plants for the identification of proteins of agronomic importance. Although by no means perfect, 2-DE coupled with mass spectrometry remains the main technology for separating and identifying complex protein mixtures in proteomic projects (GORG et al., 2004). As mentioned earlier, plant proteomics has a wide range of possibilities of investigation since this area still needs to advance considerably. We believe that plant proteomics can help improve our knowledge regarding the expression of proteins in response to specific biological questions, and can certainly provide significant contributions to the breeding programs and for the improvement of important crops.

References

AGRAWAL, G.K.; RAKWAL, R. Rice proteomics: a cornerstone for cereal food crop proteomes. Mass Spectrometry Reviews, v.25, p.1-53, 2006.

DE MOT, R; VANDERLEYDEN, J. Application of two-dimensional protein analysis of strain fingerprinting and mutant analysis of Azospirillum species. *Canadian Journal of Microbiology, v.* 35, p.960-967, 1989.

FU, B-Y; XIONG, J-H; ZHU, L-H; ZHAO, X-Q; XU, H-X; GAO, Y-M; LI Y-S, XU, J-L; LI, Z-K Identification of functional candidate genes for drought tolerance in rice. Molecular Genetics and Genomics, v.278, p.599-609, 2007.

GORANTLA, M.; BABU, P.R.; LACHAGARI, V.B.R.; REDDY, A.M.M.; WUSIRIKA, R; BENNETZEN, J.L.; REDDY, A.R. Identification of stress-responsive genes in an indica rice (*Oryza sativa* L.) using ESTs generated from drought-stressed seedlings. Journal of Experimental Botany, v. 58, p.253-265, 2007.

GORG, A.; WEISS, W.; DUNN, M.J. Current two-dimensional electrophoresis technology for proteomics. Proteomics, v. 4, p. 3665-3685, 2004.

JENSEN, O.N. Interpreting the protein language using proteomics. Nature Reviews Molecular Cell Biology, v.7, p.391-403, 2006.

JONES, A.M.E.; THOMAS, V.; TRUMAN, B.; LILLEY, K.; MANSFIELD, J.; GRANT, M. Specific changes in the Arabidopsis proteome in response to bacterial challenge: differentiating basal and R-gene mediated resistance. Phytochemistry v.65, p.1805-1816, 2004.

MEHTA, A.; BRASILEIRO, A.C.M. ; SOUZA, D.S.L.; ROMANO, E.; CAMPOS, M.; GROSSI DE SA, M.F.; SILVA, M.S.; FRANCO, O.L.; FRAGOSO, R.; BEVITORI, R.; ROCHA, T.L. Plant-pathogen interactions: what is proteomics telling us?. The FEBS Journal, v. 275, p. 3731-3746, 2008a.

MEHTA, A.; MAGALHÃES, B.S.; SOUZA, D.S.L.; VASCONCELOS, E.A.R.; SILVA, L.P.; GROSSI DE SA, M.F.; FRANCO, O.L.; COSTA, P.H.A.; ROCHA, T.L. Rooteomics: the challenge of discovering plant defense-related proteins in roots. Current Protein and Peptide Science, v. 9, p. 108-116, 2008b.

O'FARREL, P.H. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem., v. 250,p. 4007-4021, 1975

RABELLO, A. R.; QUEIROZ, P. R.; SIMOES, K. C.; HIRAGI, C. O.; LIMA, L. H. C.; OLIVEIRA, M. R. V.; MEHTA, A. DIVERSITY ANALYSIS OF Bemisia tabaci BIOTYPES BY USING RAPD, PCR-RFLP AND SEQUENCING OF THE ITS1 rDNA REGION. Genetics and Molecular Biology, v. 31, p. 585-590, 2008.

REDDY, A.R.; RAMAKRISHNA, W.; SEKHAR, A.C.; ITHAL, N.; BABU, P.R.; BONALDO, M.F.; SOARES, M.B.; BENNETZEN, J.L. Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. indica cv. Nagina 22). Genome v.45, p.204-211, 2002.

WU, C-Q; HU, H-H; ZENG, Y; LIANG, D-C; XIE, K-B; ZHANG, J-W; CHU, Z-H; XIONG, L-Z Identification of Novel Stress-responsive Transcription Factor Genes in Rice by cDNA Array Analysis. Journal of Integrative Plant Biology, v.48, p.1216-1224, 2006.

YANG, L.; ZHENG, B.; MAO, C.; QI, X.; LIU, F.; WU, P.; Analysis of transcripts that are differentially expressed in three sectors of the rice root system under water deficit. Molecular Genetics and Genomics, v.272, p.433-442, 2004.