Global Advanced Research Journal of Biotechnology Vol. 2(2) pp. 013-016, September, 2013
Available online http://garj.org/garjb/index.htm

Review

Proteomics technology as an investigative tool in plant science

Abdul Hai, Nadeem A Kizilbash* and Jamal Alruwaili

Department of Biochemistry, Faculty of Medicine and Applied Medical Sciences, Northern Border University, P.O. Box 1321, Arar-91431, Saudi Arabia.

Accepted 02, September

Large scale proteomics studies have been completed for rice (*Oryza sativa*), *Arabidopsis thaliana*, *Madicago tranculata and Nicotiana tobaccum*. There is a need to direct these studies to other plants of economic importance also. Recent developments in the methodologies of proteomics are 2-D Gel Electrophoresis for the separation of a large number of complex proteins, Mass Spectrometry for structural determination of proteins, development of protein reference maps and the use of protein microarrays. The use of tools of bioinformatics and the study of protein-protein interactions will become the future direction in Proteomics. This review discusses the current state of the art technologies being used in Plant Proteomics.

Keywords: Proteomics, Technology, Plant, Protein, Proteolytic

INTRODUCTION

The term 'Proteomics' means investigation of biological processes by the systematic analysis of a large number of expressed proteins for specific properties such as identity, quantity, activity and molecular interactions. Proteins play an important role in maintaining cellular functioning. Changes in their concentration can have direct phenotypic consequences (Alban *et al*., 2003). Several approaches have been used to identify proteins expressed in diverse organisms and organelles (Newman *et al*., 2006). More recently, efforts have been focused on a quantitative proteomics, technology that can also reveal the dynamics of biological system (Wasinger *et al*., 2013).

The understanding of the diverse role played by proteins in biological systems has increased largely as a result of the enormous amount of sequence information generated by the Human Genome Project. After completion of the Human Genome Project in 2001, approximately 25,000-30,000 genes were identified. This allowed a conservative estimate of the number of proteins present in most human cells as 25,000, although alternative splicing of genes and variations in subunit composition may increase the number of proteins even further (Cahill *et al*., 2000). Similar schemes aimed at deciphering the genomes of *E. coli* (Blattner *et al*., 1997), Yeast (Piskur and Langkjaer, 2004) and mouse (Mouse Genome Sequencing Consortium, 2002) have provided more information.

The entire genome sequencing of rice (*Oryza sativa*) (Agrawal *et al*., 2009), *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000), *Madicago tranculata* (Stanton-Geddes *et al*., 2013) and *Nicotiana tobaccum* (Renny-Byfield *et al*., 2011) has been completed but the available data does not explain how proteins function or biological processes occur. The proteome reflects the expression of molecules which

*Corresponding author E-mail: fsd707@gmail.com

directly affects cellular biochemistry because the mRNA level is not always consistent indication of protein abundance (Guo *et al*. 2008). Proteomics provides a way to study the physiological role of protein in a wider context. It can lead to answers about the role of a protein in signal transduction pathways, about post-translational modification undergone by a protein and the role of a protein in protein-protein interactions (Olsen *et al*., 2006).

Methodologies being used in Plant Proteomics

In the past 16 years, advancements in techniques such as two-dimensional gel electrophoresis, mass spectrometry, protein reference maps and protein microarrays has rendered the solution of many of the technical problems trivial. At present, the state of usage of these techniques is:

(i) 2-D Gel Electrophoresis: Improvement in 2-D electrophoresis methodology has improved reproducibility and ease-of-use. However, producing quality gels can still be technically challenging for some samples. The advantage of using protocol optimization and careful sample preparation is a reproducible proteome reference map that can be used for comparative investigations (Zychlinski and Gruissem, 2009; Thiellement *et al*., 1999; Damerval *et al*., 1989).

The protein sample for 2-D electrophoresis should be free of salts, ionic detergents, non-protein macromolecules (nucleic acids, starch, lipids) and particulate matter and present in a lyophilized or concentrated state (> 1 mg mL^{-1}). Typically, if the sample has met these requirements it can be directly re-suspended in extraction media. For samples that do not meet these criteria, further steps need to be taken. Often, a single protein precipitation step (acetone, methanol or TCA) followed by thorough washing of the pellet with cold 80% acetone is sufficient (Vienne *et al*., 1999).

(ii) Protein Reference Maps: 2-D gel electrophoresis-provided protein reference maps of different plant species have become an important tool for organizing and understanding plant proteomics. A few web sites containing 2-D gel electrophoresis data are already available (http://sphinx.rug.ac.be:8080/ppmdb/index.html; http://www.biokemi.su.se/chloroplast/, and http://www.expasy.ch/ch2d/). Reference 2-D gel electrophoresis maps can be used to follow differential protein expression and posttranslational modifications.

Some years ago, an attempt was made to map proteins found in different tissues of rice (Komatsu and Tanaka, 2005) and *Arabidopsis* (Tsugita *et al*., 1996). The proteins of different tissues were separated by 2-D Electrophoresis and a small number of abundant proteins were identified by Edman sequencing. At that time, the technology was not sufficient to obtain an analysis of the total proteomics of the different plant tissues at significant depth. Given the current level of technology, a large-scale proteomics facility with true high throughput technology might be able to obtain a more thorough characterization of the complete proteomes of the different tissues (Boschetti and Righetti, 2009).

(iii) Mass Spectrometry: Proteins of interest for biological researchers are usually present in a complex mixture of other proteins and molecules. This creates two problems: (a) the two ionization techniques used for large molecules only work well when the mixture contains roughly equal amounts of constituents. But in biological samples, different proteins tend to be present in different amounts (Issaq *et al*., 2002). If such a mixture is ionized using electrospray ionization mass spectrometry or MALDI, the more abundant species have a tendency to "drown" signals from less abundant ones; (b) the second problem is that the mass spectrum from a complex mixture is very difficult to interpret due to the overwhelming number of signals. This is made worse by the fact that enzymatic digestion of a protein gives rise to a large number of peptide products (Merchant *et al*., 2000).

Tandem MS (MS/MS) is becoming a more popular experimental method for identifying proteins. It is used to generate a set of fragments from a specific peptide ion (Mo *et al*., 2002; Negroni, 2007). The fragmentation process produces cleavage products by the rupture of peptide bonds. Because of the expected pattern of protein fragmentation, it is possible to use the fragment masses to match with a database of predicted masses. Tandem MS of whole protein ions has been investigated using electron capture and dissociation (Lin *et al*., 2003; Reid and McLuckey, 2002; Graves *et al*., 2002). This is sometimes referred to as the "top-down" approach because it involves starting with the whole mass and then pulling it apart rather than starting with proteolytic fragments and piecing the protein back together using *de novo* repeat detection. A number of different algorithmic approaches have been described to identify peptides and proteins from tandem mass spectrometry data, including peptide fragment fingerprinting, peptide *de novo* sequencing and sequence tag-based searching (Hutchens *et al*., 1993).

Peptide mass fingerprinting uses the masses of proteolytic peptides as input to a search from a database of predicted masses that would arise from digestion of a list of known proteins (Callea et al., 1999; Pappin *et al*., 1993; Mann *et al*., 1993; Henzel *et al*., 1993; James *et al*., 1993; Yates *et al*., 1993). If a protein sequence in the reference list produces a significant number of predicted masses that match the experimental values, it is likely that this protein is present in the original sample.

(iv) Protein Microarrays: Protein microarrays or "chips" are becoming more popular as miniaturized ligand-binding assays which can be used to analyze complex protein samples (Hall *et al*., 2007; Templin *et al*., 2002; Haab *et al*., 2001). They use antibodies

immobilized in a small area on a solid support such as a treated glass microscope slide. When exposed to the sample solution, each antibody captures its target protein. This technique allows large scale and high throughput analysis (Lopez *et al*., 2003). Protein microarrays effectively allow quantitation of several hundred to several thousand analytes with one system (MacBeath, 2002). Antibody microarrays are the most accessible medium used in proteomics. In one of the earlier papers reporting the use of protein microarrays, 115 antibodies or antigens were immobilized using robotics and investigated with the corresponding ligands in mixtures of varying but known concentrations (Wingren *et al*., 2003). The disadvantages of using antibodies include the large molecular weight, and for polyclonal antibodies, a lack of specificity. Commercial companies are, at present, developing methods for overcoming these problems using antibody fragments or phage technology, the latter using phage (bacterial viruses) which have been genetically modified to express immunoglobulin fragments on their surface (Madoz-Gurpide *et al*., 2001; Wingren *et al*., 2003).

Future Direction

Like the Human Proteome Project (Legrain *et al*., 2011), that was designed to map the entire set of human proteins, a similar effort is needed for plants such as rice (*Oryza sativa*), *Arabidopsis thaliana*, *Madicago tranculata and Nicotiana tobaccum*. Given the lack of protein-level information present for most of the protein-coding genes identified so far in plants, a systematic effort using the tools of bioinformatics and the investigation of protein-protein interactions is necessary to achieve this goal.

REFERENCES

Alban A, David SO, Bjorkesten L, Andersson C, Sloge E, Lewis S and Currie I (2003). A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics, 3: 36–44.

Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997). The complete genome sequence of *Escherichia coli* K-12. Science 277(5331): 1453-62.

Boschetti E and Righetti PG (2009). The art of observing rare protein species in proteomes with peptide ligand libraries. Proteomics 9(6): 1492-1510.

Cahill DJ, Nordhoff E, O'Brien J, Klose J, Eickhoff H and Lehrach H (2000). In: S. Pennington and M. Dunn (Eds.) Bridging Genomics and Proteomics (pp.1-17). Oxford, England: BIOS scientific publisher.

Callea L, Arese M, Orlandini A, Bargnani C, Priori A and Bussolino F (1999). Platelet activating factor is elevated in cerebral spinal fluid and plasma of patients with relapsing-remitting multiple sclerosis. J. Neuroimmunol. 94: 212–21.

Damerval C, Zivy M, Granier F and de Vienne D (1989). Two-dimensional

electrophoresis in plant biology. Adv. Electrophoresis 2: 263–340.

Graves PR and Haystead T (2002). Molecular biologist's guide to proteomics. Microbiol Mol Biol Rev. 66: 39–63.

Guo Y, Xiao P, Lei S, Deng F, Guishan G, Yaozhong X, Xiangding L, Li CL, Yuan S W, Jiang CH, Jingyun LT, Zhu XX, Liang S and Deng H (2008). How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. Acta Biochim Biophys 40(5): 426-436.

Haab BB, Dunham MJ and Brown PO (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol. 2(4): 1-4.

Hall DA, Ptacek J and Snyder M (2007). Protein Microarray Technology. Mech Ageing Dev. 128(1): 161–167.

Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc. Natl. Acad. Sci. U.S.A. 90 (11): 5011–5.

Hutchens TW and Yip TT (1993). New desoprtion strategies for the mass-spectrometric analysis of macromolecules. Rapid Commun Mass Spectrom. 7: 576–80.

James P, Quadroni M, Carafoli E and Gonnet G (1993). Protein identification by mass profile fingerprinting. Biochem. Biophys. Res. Commun. 195 (1): 58–64.

Komatsu S and Tanaka N (2005). Rice proteome analysis: a step toward functional analysis of the rice genome. Proteomics. 5(4): 938-49.

Legrain P, Aebersold R, Archakov A, Bairoch A, Bala K, Beretta L, Bergeron J, Borchers CH, Corthals GL, Costello CE, Deutsch EW, Domon B, Hancock W, He F, Hochstrasser D, Marko-Varga G, Salekdeh GH, Sechi S, Snyder M, Srivastava S, Uhlén M, Wu CH, Yamamoto T, Paik YK and Omenn GS (2011). Molecular and Cellular Proteomics, 10, M111.009993.

Lin D, Tabb DL, Yates JR III (2003). Large-scale protein identification using mass spectrometry. Bioch. Biophys. Acta 1646: 1–10.

Lopez MF and Pluskal MG (2003). Protein micro- and macroarrays: digitizing the proteome. J Chromatogr B. 787:19–27.

MacBeath G. Protein microarrays and proteomics (2002). Nat Genet Suppl. 32, 526–32.

Madoz-Gurpide J, Wang H, Misek DE, Brichory F and Hanash SM (2001). Protein based microarrays: a tool for probing the proteome of cancer cells and tissues. Proteomics. 1: 1279-387.

Mann M, Højrup P and Roepstorff P (1993). Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol. Mass Spectrom. 22 (6): 338–45.

Merchant M and Weinberger SJ (2000). Recent developments in surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. Electrophoresis. 21: 1164–7.

Mo W and Karger BL (2002). Analytical aspects of mass spectrometry and proteomics. Curr Opin Chem Biol. 6: 666–7.

Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520-562.

Negroni L (2007). Protein identification using nano liquid chromatography-tandem mass spectrometry. Methods Mol Biol. 355: 235-248.

Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL and Weissman JS (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature. 2006 441(7095): 840-846.

Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P and Mann M (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127(3): 635-648.

Pappin DJ, Hojrup P and Bleasby AJ (1993). Rapid identification of proteins by peptide-mass fingerprinting. Curr. Biol. 3 (6): 327–332.

Piskur J and Langkjaer RB (2004). Yeast genome sequencing: the power of comparative genomics. Mol Microbiol. 53(2): 381-389.

Reid GE and McLuckey SA (2002). 'Top down' protein characterization via tandem mass spectrometry. J Mass Spectrom. 37(7): 663-75.

Renny-Byfield S, Chester M, Kovar A, Le Comber SC, Grandbastien MA, Deloger M, Nichols RA, Macas J, Novák P, Chase MW and Leitch AR (2011). Next Generation Sequencing Reveals Genome Downsizing in Allotetraploid *Nicotiana tabacum* Predominantly through the Elimination of Paternally Derived Repetitive DNAs. Mol Biol Evol. 28 (10): 2843-2854.

Stanton-Geddes J, Paape T, Epstein B, Briskine R, Yoder J, Mudge J, Bharti AK, Farmer AD, Zhou P, Denny R, May GD, Erlandson S,

Yakub M, Sugawara M, Sadowsky MJ, Young ND and Tiffin P (2013). Candidate Genes and Genetic Architecture of Symbiotic and Agronomic Traits Revealed by Whole-Genome, Sequence-Based Association Genetics in *Medicago truncatula*. PLoS One, 8(5): e65688.

Templin MF and Stoll D, Schrenk M, Traub PC, Vohringer CF and Joos TO (2002) Protein microarray technology. Drug Discovery Today. 7: 815–22.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796-815.

Thiellement H, Bahrman N, Damerval C, Plomion C, Rossignol M, Santoni V, de Vienne D and Zivy M (1999). Proteomics for genetic and physiological studies in plants. Electrophoresis 20, 2013–202

Tsugita A, Kamo M, Kawakami T and Ohki Y (1996). Two-dimensional electrophoresis of plant proteins and standardization of gel patterns. Electrophoresis 17: 855-865.

Wasinger VC, Zeng M and Yau Y (2013). Current Status and Advances in Quantitative Proteomic Mass Spectrometry. *International Journal of Proteomics* Volume, Article ID 180605, 12 pages.

Wingren C, Ingvarsson J, Lindstedt M and Borrebaeck C (2003). Recombinant antibody microarrays- a viable option? Nat Biotechnol. 21, 223.

Yates JR, Speicher S, Griffin and PR, Hunkapiller T (1993). Peptide mass maps: a highly informative approach to protein identification. Anal. Biochem. 214 (2): 397–408.

Zychlinski AV and Gruissem W (2009). Preparation and Analysis of Plant and Plastid Proteomes by 2DE. Methods Mol Biol. 519: 205-220.