A perspective on proteomics: current applications, challenges and potential uses

[1]Muturi, P.W♠., [1]Mwololo, J.K., [1]Munyiri, S.W., [1]Rubaihayo P., [2]Munyua, J.K., [3]Mgonja M. [3]Manyasa E, [4]Kiarie N

[1]Makerere University, Crop Science Department P.O. Box 7062 Kampala, Uganda;
[2]University of Nairobi, Biochemistry Department, P.O. Box 30197, 00100 Nairobi, Kenya.
[3] International Crops Research Institute for Semi-Arid Tropics, P.O. Box 39063-00623, Nairobi, Kenya
[4]University of Nairobi, Plant Science and Crop Protection Department P.O. Box 29053-00625 Nairobi, Kenya

ABSTRACT

Biological sciences are experiencing an ongoing information revolution. Proteome-wide functional classification using bioinformatics approaches is becoming an important method for revealing unknown protein functions. Most successful computational approaches for protein function prediction integrate multiple genomics and proteomics data sources to make inferences about the function of unknown proteins. Research into gene expression and proteomics enable scientists to decipher the functions of genes and their protein products, and to get a clearer picture of the complex regulatory networks that control fundamental biological processes. The global study of cellular proteins by proteomics may be able to provide the complete picture. Use of proteins to study gene function and genetic information is possibly the most reliable method but costly and labour intensive. Analysis of gene-expression patterns is no less powerful concept than proteomics when it comes to identification of the characteristics of signalling pathways or disease states. This review discusses current applications of proteomics, challenges and potential uses.

Key words: Proteomics, functional genomics, physiological processes, gene expression

INTRODUCTION

Proteomics refers to the analysis of protein profiles. The amino acid sequence of a protein determines the protein's three-dimensional structure and function (Banks et al., 2000). One of the goals of functional genomics and its approaches is to understand what the function of each of the genes and their protein products is (Tytus et al., 2006). Systemic deletion approaches can demonstrate gene function by demonstrating functional deficiencies in an organism when the gene is removed (Banks et al., 2000). It is worth noting that function is a protean concept. DNA and RNA fulfill the tasks of storage, transfer and processing of genetic information contained in the genome of living organisms (Tytus et al., 2006). Proteins form complex cellular machinery for the realization of this genetic program resulting in the phenotype in dependency and in response to changing environment conditions (Nakanishi et al., 2001).

The identification of candidate genes influencing any important trait can be approached through analysis of complementary DNA (cDNA), copies of messenger RNA (mRNA). These mRNAs represent only a small percentage of the total genome (about 1-3% in eukaryotes) (Banks et al., 2000; Tytus et al., 2006). However, they do contain valuable information on gene activity since they correspond to the proteins expressed in a specific tissue and responsible for the identity of that tissue. Following extensive development in mammals and bacteria, there has been an increasing analysis of protein profiles in plants. The number of studies includes those on legume root nodules associated with nitrogen fixing bacteria among others (Panter et al, 2000). This review discusses current applications of proteomics, challenges and potential uses.

Current application of proteomics : The availability of sequence information related to proteins provide an excellent opportunity to conduct bioinformatics studies to discover and understand unexpected relationships between living organisms (Tytus et al., 2006). For instance proteins of unknown function can be classified based on their pattern of sequence similarity to sequences of other organisms (Gutiérrez et al., 2004). Proteomics has been extensively used in understanding and treatment of molecular diseases in humans (Rossignol, 2000; Tytus et al., 2006). Efforts in

proteomic technologies address the key limits of the current technologies: the narrow spectrum and quantification of the protein properties that can be monitored simultaneously (Pandey and Mann, 2000).

Full realization of the information encoded in genome sequences requires knowledge of the three-dimensional (3-D) structures of gene products, since it is at this level that gene function is expressed (Rossignol, 2001;Tytus *et al.,* 2006). Protein 3-D structure has traditionally provided the basis for understanding functions that have already been determined biochemically and for applications in medicine and biotechnology such as protein engineering and structurebased drug design (Seung *et al.,* 2006). Now, however, with increased throughput, it also offers a route to the discovery of function for the many gene products that are currently uncharacterized (Gutiérrez *et al.,* 2004). The use of amino acid derived data provides perhaps the most accurate functional estimation. This is mainly because phenotype of an organism is largely expressed by a combination of enzymes that are proteins in nature (Pandey and Mann, 2000; Tytus *et al.,* 2006). The use of this technique has provided evidence of a particular expression pattern within a specific cell or tissue and has made a major contribution in human and microbial functional genomics (Seung *et al.,* 2006).

Using protein family information to predict gene function is more reliable than using sequence comparison alone (Rossignol, 2001). One can use structure or function-based protein families when available to complement sequence-based family for additional function information (Tytus *et al.,* 2006). Phylogenetic analysis using proteins has been used in comparative genomics, gene function prediction and inference of lateral gene transfer among other functions (Banks *et al.,* 2000). Proteins are highly conserved thus can be used to study relationship across species, proteins unlock information enclosed in nucleotide and proteins are the ultimate expression of a gene thus useful in studying function (Rossignol, 2001; Tytus *et al.,* 2006).

Two-dimensional gel electrophoresis allows the separation of cellular proteins on a polymer gel according to their molecular weight in one dimension and according to their isoelectric point in the second dimension (Banks *et al.,* 2000; Seung *et al.,* 2006). This technique allows the quantification of sets of cellular proteins and similarly the gene expression monitoring and comparative studies provide invaluable information about the cellular function of proteins whose elementary function can

either be known or unknown (Pandey and Mann, 2000; Tytus *et al.,* 2006)). Another set of proteomic technologies allows the identification of interactions between proteins and between proteins and DNA domains (Gutiérrez *et al.,* 2004). These kinds of interactions suggest the involvement of the corresponding proteins in the regulation of signal transduction and transcription (Tytus *et al.,* 2006). Efforts in proteomic technologies today address the key limits of the current technologies: the narrow spectrum and quantification of the protein properties that can be monitored simultaneously (Seung *et al.,* 2006).

The development of automated methods for the annotation of predicted gene products (proteins) with functional categories is becoming increasingly important, in order to present genome sequences and genome annotations to biologists in a useful way (Tytus *et al.,* 2006). Thus, many systems to perform protein functional annotation have been developed that employ various sources of protein information as features, including protein functional sites, sequence similarity and gene expression patterns (Pandey and Mann, 2000; Seung *et al.,* 2006).

Given the number of proteins that can be produced by individual organisms, it seems that proteomics may allow greater understanding of the complexity of life and the process of evolution than the study of the genetic code alone (Banks *et al.,* 2000). Proteomics doesn't only reveal information about life's complexity for it also provides insight into the vibrancy of cells and their preparedness to react (Seung *et al.,* 2006). Cells and tissues respond to signals and changes in their environment and changes in the proteome must mirror that (Pandey and Mann, 2000; Tytus *et al.,* 2006). In fact, early changes in the health of a tissue may be detectable by changes at the proteomic level. Researchers are currently taking advantage of measurable changes in protein profiles to assess new diseases emerging (Banks *et al.,* 2000; Gutiérrez *et al.,* 2004). Moreover, the difference in profiles is robust enough to be used as a predictive diagnostic tool.

Challenges in utilization of proteomics : Certain disadvantages are limiting the use of proteomics. Proteins are dynamic and interacting molecules, and their changeability can make proteomic snapshots difficult. There is the need for a more sensitive analytical system and the absence of an effective method for large-scale data comparison (Pandey and Mann, 2000). Very closely related proteins may not guarantee a functional relationship (Dumwell *et al.,* 2001). The false positive rate of motif assignment is high due to high

probability of matching short motifs in unrelated proteins by chance (Rossignol, 2001).

Proteins possess different levels of energy for instance hydrogen bonds, hydrophobic effects, vanderwaal forces and electrostatic forces thus difficult to deal with proteins. In modeling proteins structure the different levels of energy in protein molecule influence accuracy that can be obtained yet they are essential in building tertiary and quaternary structure of protein (Banks *et al.,* 2000). Unlike DNA sequencing, protein sequencing is a relatively costly and laborious process. The other challenge is that there are a few protein sequences available and if available they are either too short or highly conserved thus difficult to study variation (Seung *et al.,* 2006; Tytus *et al.,* 2006).

Within the proteome, the many observed layers of complexity begin with an RNA processing mechanism called alternative splicing in which a single gene can produce multiple versions of a protein (Rossignol, 2001). An example is the production of neurexins in mammals whereby three genes give rise to over 1,000 distinct proteins within the mammalian brain (Seung *et al.,* 2006). Post-translational modifications are also another source of protein variation. More than 200 different types of post-translational modifications are known and it is predicted that, for each gene in eukaryotes three different modified proteins with different functions are produced (Banks *et al.,* 2000; Rossignol, 2001). Characterizing the biochemical and cellular functions of each protein and the analysis of protein regulation and its relation to other regulatory networks also poses a challenge (Tytus *et al.,* 2006).

CONCLUSION AND FUTURE PROSPECTS:
Proteomics has emerged as an indispensable methodology for large-scale protein analysis in functional genomics. Proteomics will remain to be one of the fastest growing areas in research. Proteomics is a useful and powerful tool for investigating protein changes induced by various conditions. The global-scale analysis of proteins is expected to yield more direct understanding of function and regulation than analysis of genes. To meet the current challenges of food insecurity, genes and proteins that control crop architecture and/or stress resistance in a wide range of

environments will need to be identified to facilitate the biological improvement of crop productivity, However there is need of integration of genomics, trancriptomics and proteomics to facilitate understanding of normal function, disease, and development. As a result, systems biology approaches will continue to detect connections between broad cellular functions and pathways that were neither apparent nor predictable despite decades of biochemical and genetic analysis of the biological system in question.

REFERENCES

Banks, R.E., Dunn, M.J., Hochstrasser, D.F., Sanchez, J.C., Blackstock, W., Pappin, D.J. and Selby, P.J (2000). Proteomics: New perspectives, new biomedical opportunities. Lancet 356:1749-1756

Brendel, V. and Zhu, W (2002). Computational modeling of gene structure in Arabidopsis thaliana. Plant Mol. Biol. 48:49–58

Dunwell, J.M., Maria, A.M and Raul, H (2001). Transcriptome analysis and crop improvement. Biological Research. 34; 3-4

Dutt, M.J and Lee, K.H (2000). Proteome analysis. Curr Opin Biotechnol 11:176-179

Gutiérrez, R.A., Green, P.J., Keegstra, K. and Ohlrogge, J.B (2004). Phylogenetic profiling of the *Arabidopsis thaliana* proteome: what proteins distinguish plants from other organisms? Genome Biol.; 5(8):R53.

Jiang, N., Bao, Z., Zhang, X., Eddy, S.R and Wessler, S.R (2004). Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–73

Pandey, A and Mann, M (2000). Proteomics to study genes and genomes. Nature 405:837-846

Panter, S., Thomson, R., Bruxelles, D.E. G., Laver, D., Trevaskis, B. and Udvardi, M (2000). Identification with proteomics of novel proteins associated with the peribacteroid membrane of soybean root nodules. Mol Plant Microbe Interact 13:325-333

Rossignol, M (2001). Analysis of the plant proteome. Current Opinion. Biotechnol 12:131-134

Seung, Y. R., Julie D and Dong, X (2006). Bioinformatics and its Applications in Plant Biology. Annual Review of Plant Biology. 57: 335-360

Tytus, B., Gérald, G., Eli, K.A and Paul, R. J (2006). Integrating Cytomics and Proteomics. Molecular and Cellular Proteomic*s 5:*2-13, 2006.