Current initiatives in proteomics research: The plant perspective

Himanshu Dubey and Anil Grover*

Department of Plant Molecular Biology, University of Delhi South Campus, Benito Juarez Road, Dhaula Kuan, New Delhi 110 021, India

The recent upsurge in structural genomics is leading to the accumulation of a huge wealth of literature on nucleotide sequences. After the nucleotide sequence of a given stretch of DNA is obtained (by manual or robotic methods), the next step is to use some software program to distinguish the possible open reading frames in the thick of sequences. However, the acid test, whether the new sequence corresponds to any functionality in terms of transcription and translation is to identify the protein which it encodes. Functional genomics and proteomics are the buzzwords in modern-day genomics. The science of proteomics is a possible approach to relate the skeletal nucleotide sequence information to functional attributes of the cell. The identification and isolation of novel genes with potential biotechnological applications warrant that genomics and proteomics must go hand in hand. Three major steps in proteome analysis are the separation of complex protein mixtures by two-dimensional protein gel electrophoresis (2D), characterization of the separated proteins by mass spectrometer (MS) and database searching. The power of 2D is such that it allows even minor changes in gene expression caused by internal or external cues to be effectively scored. Most proteins resolved by 2D have high purity, which can facilitate their identification by MS. In recent years, methods for automated proteomics based on incorporation of new ideas in both hardware and software development have been optimized to a great deal. We discuss the progress and applications of the proteomics science, with special reference to plants.

DURING the last decade, major advances have been made in plant genetic engineering[1]. The methods for stable genetic transformation as well as regulation of introduced trans-genes have been optimized to a great extent. The widespread progress of plant genetic engineering and biotechnology research in recent times is mainly limited by the non-availability of the agronomically-important target genes[2,3]. It is therefore relevant to look for newer approaches that can lead to the identification, isolation and cloning of such genes. This article takes a look at current developments in 'proteomics' science. It particularly emphasizes the application of proteomics in isolation of important novel plant genes.

*For correspondence. (e-mail: grover_anil@hotmail.com)

Proteomics research

'Genomics' is the most recent 'happening' science in contemporary biology[4,5]. Complete genomes of more than 30 organisms (e.g. *Escherichia coli, Bacillus subtilis, Synechocystis* sp., *Mycobacterium tuberculosis, Treponema pallidum, Borrelia burgdorferi, Deinococcus radiodurans, Aquifex aeolicus, Saccharomyces cerevisiae* and *Caenorhabiditis elegans*) have already been sequenced and the genomes of another 100 or so organisms, including that of human, are being sequenced at a rapid pace. Among plants, the genome of *Arabidopsis thaliana* has already been completely sequenced and the genome of *Oryza sativa* is a target for complete nucleotide sequencing within the next 3–4 years. Apart from these, there is a great deal of progress in genomics of maize, sorghum, sugarcane, barley, cotton, tomato, soybean, tobacco and *Chlamydomonas reinhardtii* (for further details on plant genomics research the reader may refer to refs 5–10).

In this exciting age of nucleotide sequencing, it is important not to forget that, in reality, genome projects simply transfer digital information from DNA to computer files and this is a long way from providing an understanding of function. Knowing the genome sequence and even the location of all genes of an organism is the anatomical description of its genome. It is therefore important that the nucleotide sequence is further translated into functionality, meaning that genes identified through such projects are ascribed roles in terms of encoded proteins. 'Functional genomics' is the science of understanding how the genome works through a control on the expression of genes.

However, the understanding of the biological functions of the novel genes is a more difficult proposition than obtaining just the sequence. This is because of the fact that the existing amount of information on amino acid sequences of known proteins in the database does not match the wealth of information on nucleotide sequences being generated through genome projects[4,5,11]. Already, it is seen that nearly 50% of the nucleotide sequences coming from the genomics research do not

match in homology with any known proteins[4,5,11]. As has been the case for a number of prokaryotic and eukaryotic species whose genomic DNA sequences have been completed, many open reading frames (ORFs) in *Arabidopsis* and rice genome, when completed, will encode proteins of unknown functions.

A possible solution in this regard is the use of 'microarray chip' technology[4]. The genome-wide analysis of mRNA expression by microarray approach is providing important clues about expression patterns and thus the functions of gene products[4,8]. However, for a substantial number of proteins, there may be only a loose correlation between mRNA and protein levels[11,12]. In addition, the functions of proteins depend considerably on post-translational modifications and interactions with other proteins, processes that cannot be deduced from nucleic acid microarray data. The analysis of proteins is the most direct approach to define the gene function. Therefore, efficient approaches for identifying proteins, for determining profiles of protein expression in different tissues and under different conditions, for identifying post-translational modifications of proteins in response to different stimuli and for characterizing protein interactions are critical for understanding biological processes in the post-genome era[11,13]. If a new ORF found in the genomics programme is strongly correlated to a specific developmental stage- or inducer-dependent protein, it would lead to unveiling the function of that ORF. Therefore, protein analysis or proteomics is turning out to be a major international subject of research after the structural genomics wave.

Proteomics means investigation of biological processes by the systematic analysis of a large number of expressed proteins for specific properties such as their identity, quantity, activity and molecular interactions[11,14]. Over the past few years, significant progress has been made towards developing a mature technology for the identification and cataloguing of the proteins expressed in a cell or tissue (so-called 'descriptive proteomics'). More recently, efforts have been focused on a 'quantitative proteomics' technology that can also capture the dynamics of the biological systems[15].

Technical inputs in proteomics research

A critical requirement in proteomics research is high quality separation of cellular proteins. Such analysis of proteins by electrophoresis has been a subject of active research for a long time. The major historical milestones of the protein gel electrophoresis technique are presented in Table 1. O'Farrell[16] made a path-breaking advancement in electrophoretic protein separation methods by combining iso-electric focusing (IEF) and sodium dodecyl sulphate (SDS) gel electrophoresis, resulting in the powerful two-dimensional gel electrophoresis (2D) technique. At the time of its inception, this technique resolved some 1100 different proteins of *E. coli*. Several innovations made in the basic 2D technique as shown below and further in Box 1, have made this technique suitable for a range of different applications. This method has been optimized for the separation of both soluble as well as membrane protein fractions[17]. The glycoproteins as well have come under the analytical power of 2D[18]. 2D-separated proteins can be subjected to analysis of amino acid composition, immunological characterization and peptide mapping. The amino acid sequence of 2D-separated proteins or of peptides generated and purified from 2D-separated proteins can be directly determined[19].

Table 1. Selective milestones in development of the protein gel electrophoresis technique

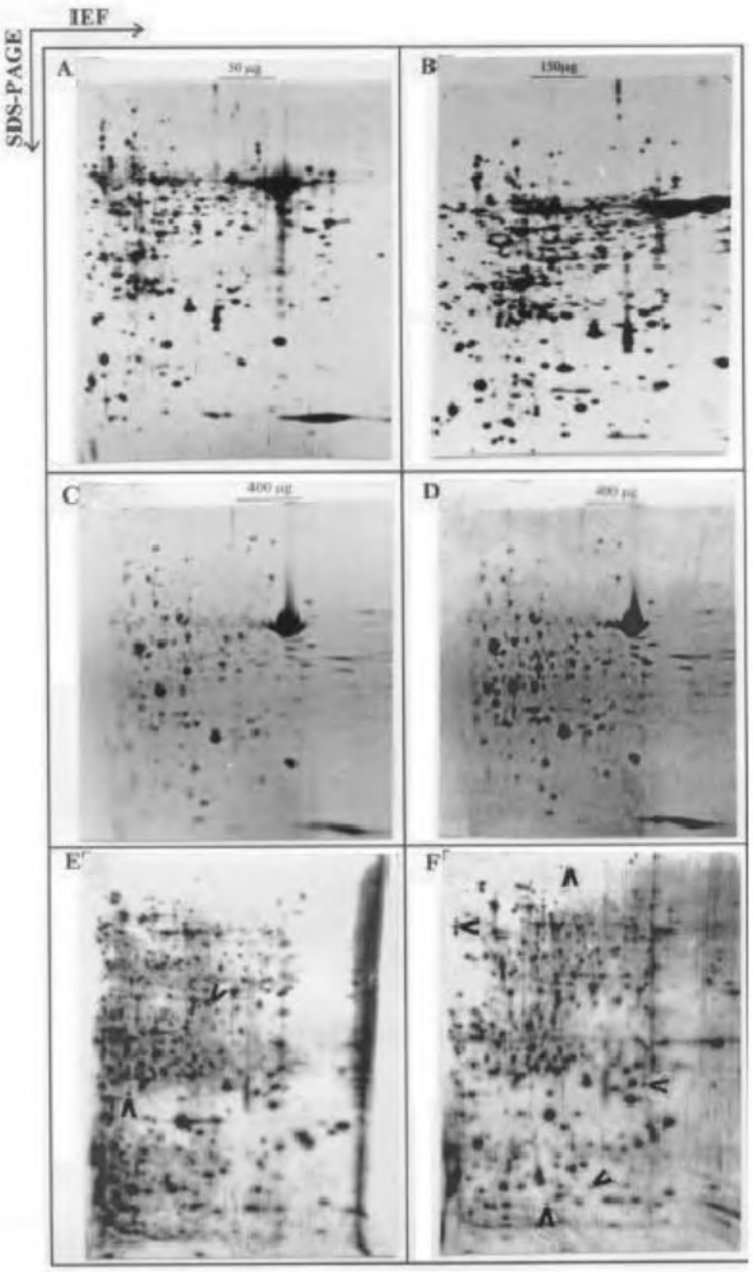
Advances in technique	Reference
Free-boundary electrophoresis	66
Introduction of acrylamide	67
Iso-electric focusing (IEF)	68
Disc electrophoresis	69
Introduction of SDS	70
Gradient gel	71
SDS coupled to discontinuous buffer system	72
2D (IEF + SDS-gel) analysis	16
Silver staining	73
Immobilized pH gradients	74
Computer digitization of protein 2D map	75

The next important development in proteomics research is the combination of 2D with mass spectrometry (MS) for the analysis of the separated proteins[20]. MS has essentially replaced the classical technique of Edman degradation because not only is it far more sensitive, it can also deal with protein mixtures and offers much higher throughput[11]. Typically, the MS technique consists of a source to generate ions from the sample and an analyser to separate and detect these ions according to their mass. For proteomics, two sources are widely used, namely matrix-assisted laser desorption-ionization (MALDI) and electrospray ionization (ESI) and analyzers range from the simple (e.g. time of flight or TOF) to complex (e.g. Fourier transform ion cyclotron resonance or FTICR). Simple MS such as MALDI-TOF measures only the mass. However, 'tandem mass spectrometry' also allows the amino acid sequence to be determined (Figure 1; for more technical details on terms such as ESI, MALDI, nESI, PMF, PSD, PI, TM, TQ and QTOF associated with MS, see refs 11 and 20).

The MS quantitation of the masses of a few tryptic fragments from an unknown protein, followed by the use of algorithms to compare the observed peptide masses against those predicted for the theoretical tryptic fragments of all expressed sequences (from database

Box 1.

The classical 2-dimensional protein gel electrophoresis (2D) work of O'Farrell[16] was based on pH gradients generated by carrier ampholytes (CA). In instances where 2D spots are to be used for eluting sufficient amounts of the proteins, for raising antisera or for detection of the identity, increased loading of the sample is desirable. The panels A and B (see figure) represent consequences of over-loading of the sample proteins in a 2D gel made using CA. Rice-seedling proteins were isolated as per the method of Suzuki *et al.*[86]. The amounts of the soluble proteins were quantified according to Bradford[87]. The 2D gel electrophoresis was carried out essentially as per the procedure of Suzuki *et al.*[86]. The IEF gel consisted of 2:8:1 volumes of 3.5–10:5.0–8.0:8.0–10.5 ampholines (Sigma, USA). The second dimension gel consisted of 12.5% polyacrylamide concentration. Following the completion of the run, the gel was silver-stained. It is clear from comparison of these panels that increased protein loading (panel A-50 µg and panel B-150 µg) causes the pH gradient to move towards the cathode (cathodic drift) and flattens in the centre (plateau phenomenon). Cathodic drift is also known to result due to prolonged focusing durations[19]. By using immobilized pH gradient gels (IPG), cathodic drift is eliminated and reproducible IEF patterns are obtained[19].



The resolved proteins on 2D gel can be detected by chromophoric staining with Coomassie brilliant blue (CBB) or silver nitrate, autoradiography or phosphour-imaging of radiolabelled proteins or staining with fluorescent dyes. CBB staining is easy to perform and shows good reproducibility but is not very sensitive and only allows detection of major components within a protein sample (typically, not more than 200–300 spots). Silver staining is up to 100-fold sensitive, which makes it ideal for the detection of trace components within a protein sample (the detection limit is 0.1 ng protein) and is still the method of choice for analytical gels (compare panel A which corresponds to 50 µg protein and silver stained with panel C which corresponds to 400 µg protein of same protein sample and CBB-stained).

The major objectives in computer-aided image analysis are data acquisition, background reduction, removal of streaks, spot detection and quantitation, pattern matching, database construction and data analysis[88]. Images are captured either by scanning the silver- or CBB-stained gels and/or the exposed X-ray films using flat-bed scanners, CCD cameras or laser densitometers. Panel D represents CCD camera-based imaging of the panel C. The images are then edited and matched using 2D image analysis software.

The usefulness of the proteomic approach is highly dependent on experimental design[63]. One common approach is to compare 2D maps of samples derived from non-induced (control) with induced (stressed) tissues. Panel E represents a non-induced rice shoot sample and panel F an induced rice shoot sample (by subjecting to 6 h of anoxia stress, see Hossain *et al.*[89] for details on the stress treatment). Induced proteins (shown by arrowheads in panel F) could be important in the cellular response to a given environmental change. Repressed proteins (shown by arrowheads in panel E) may no longer be needed in adapting to a new environment. Induced proteins may be the result of increased transcription or activation, while those repressed may be due to decreased transcription, activated degradation or inactivation through post-translational modifications.

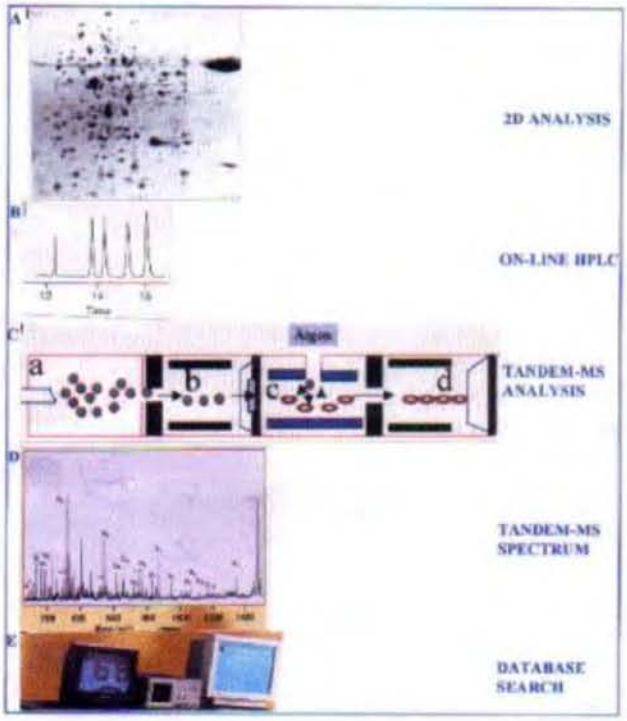


Figure 1. Steps involved in proteome analysis by tandem mass spectrometry. **A**, Proteins are separated by 2D and stained spots are excised and subjected to in-gel digestion with trypsin; **B**, The resulting peptides are separated by on-line high-performance liquid chromatography (HPLC); **C**, The eluting peptides are subjected to tandem mass spectrometry in which two stages of mass analysis are linked in series. The peptides are ionized by electrospray ionization (a), then delivered to first MS in which peptides are identified based on their mass to charge (m/z) ratio (b), further selected peptide is fragmented by collision with an inert gas such as argon (c), and finally the second MS analysis separates the resulting peptide fragments (d); **D**, MS spectrum of the peptide gives valuable information on amino acid sequence; **E**, Amino acid sequence information helps in homology searching and cloning or database identification of the corresponding gene.

which have the genome sequences), suffices for exact protein identification[11,21]. This process known as 'peptide mass fingerprinting', is a powerful method for protein identification and expression pattern analysis[22]. Recently developed techniques allow automation of in-gel tryptic digestion of all the proteins in the 2D gel, followed by their transfer to a membrane that can be scanned by MS to obtain diagnostic peptide masses for peptide mass fingerprinting[23]. However, until the complete genome sequence is not available for peptide mass fingerprinting, the amino acid sequencing approach will continue to be of great help for protein identification. For faster automated analysis of a large number of samples, tandem mass spectrometric sequencing is the method of choice. Amino acid sequence information provided by MS analysis can allow homology searching, cloning and database identification of the corresponding gene[13,24,25]. Li and Assmann[13] recently subjected AAPK (ABA-activated and Ca^{++}-independent protein kinase) excised from 2D to tandem MS for amino acid sequence analysis. Peptide sequence obtained from AAPK turned out to be highly conserved in PKABA1 class of protein kinases. This information enabled cloning of the AAPK cDNA. Truly, the synthesis of 2D + MS is the backbone of the present-day proteomics science.

Further support for proteomics has been lent by the emerging computer technologies[26,27]. Identifying proteins by mass requires access to a protein sequence database. The most commonly used databases are SWISS-PROT, TrEMBL and non-reduntant collection of protein sequences at the US National Centre for Biotechnology Information (NCBI). SWISS-PROT is an annotated collection of protein sequences. The NCBI database contains translated protein sequences from the entire collection of DNA sequences kept at GeneBank. Some of the important databases are as follows: (a) SWISS-2DPAGE for protein identification; (b) NCBI/BLAST and SWISS-PROT which are sequence databases; (c) SWISS-MODEL for three-dimensional structure; (d) PROSITE (e.g. PIR, SWISS-PROT) for domain structure, and finally (e) GenBank and EMBL which are DNA data banks. The science, of bioinformatics is a cardinal part of the present-day proteomics science, as development of sophisticated software for an efficient analysis and storage of data with partially automated comparison of multiple 2D gels is needed in scaling proteomics to meet challenges put by the genomics research.

The protein interactions can be analysed directly through proteomics science by performing co-precipitation studies with a 'bait' protein followed by mass spectrometric 'read-out' of the bound proteins. Proteomics techniques aimed at identifying protein–protein interactions have been used successfully to characterize multiprotein complexes such as spliceosome[28], nuclear pore complex[29] and the transient complexes in cell signalling[30]. The architecture of the nuclear pore complex has been revealed to an unprecedented detail using the proteomics approach[29]. Nuclear pore complex is the gateway that regulates two-way traffic between the nucleus and the rest of the cell. Using MS technique, all the detectable polypeptides in the complete nuclear pore complex of yeast have been identified. This work has revealed that around 30 different

proteins constitute the nuclear protein complex, and it has been possible to localize each protein within the complex by immuno-electron microscopy and determine how much of each was present. Indeed, 'interaction proteomics' is the answer to large-scale, unbiased exploration of complexes, cellular structures and pathways.

The post-translational modification of proteins is a key regulatory event in many cellular processes, including signalling, targeting and metabolism. Current developments in proteomics enable the global analysis of post-translationally modified proteins. The MS has been employed to characterize function-critical post-translational modifications, including phosphorylation and glycosylation[31]. The feasibility of modification-specific proteomics is shown by the recent study of phosphopeptides[11].

Based on the above account, it can be inferred that the current scope of proteomics is much broader than was indicated by its classical definition that included only the 2D-based analysis[11]. It includes protein identification, study of post-translational modifications and protein–protein interactions and the determination of function. Several newer methods of protein analysis such as affinity purification, antibody usage, yeast two-hybrid system, phage display, etc. have been combined with 2D to effectively address to these objectives[11].

Current applications of proteomics in varied biological systems

Proteomics has been aptly called the 'science in preparation for the new millennium', due to rapid advances achieved in its automation, combinatorial chemistry and high throughput screening[32,33]. On account of its enormous potential, proteomics can be further divided into 'expression proteomics' (study of global changes in protein expression) and 'cell-map proteomics' (systemic study of protein–protein interactions). Proteomics is proving an indispensable tool for examining alteration in the protein profiles caused due to gene mutations, introduction or silencing of genes or in response to various stress stimuli in a relatively fast, sensitive and reproducible way. This science is becoming important for generation of information on physiological (e.g. regulatory behaviour and function), biochemical (e.g. metabolic and structural data), genetic (e.g. gene mapping and the assigning of the structural genes to the 2D gel map) and architectural (e.g. location of the proteins in the cell) aspects. Proteomics-based approach is proving important for characterization of individuals or lines, estimation of genetic variability within and between different populations, establishment of genetic distances to be used in the phylogenetic studies and characterization of mutants with localization of genes encoding revealed proteins[9]. Selective applications of proteomics in animal and microbial systems include discovery of target molecules, designing/discovery of novel biomolecules and proteins (pharmaceutical, industrial and environmental applications), finding high-value peptides/proteins, antibodies, vaccines, enzymes, therapeutic peptides, drug discovery and biomolecular engineering[34–36].

In view of the anticipated role of proteomics in elucidating the function of the genes that are (or will be) sequenced in the near future, there are on-going attempts to establish a comprehensive protein database for a wide group of subjects by several public and private firms. The governments of several countries have allocated massive funding to proteomics in recent years. For instance, Japan has taken a strong initiative to substantially boost research on proteomics in the government budget request for the financial year 2000 (ref. 37).

Plant proteomics research

Proteomics is becoming a necessity in plant biology for deciphering the function and the role of genes in the ongoing plant genome sequencing projects. The applications of proteomics can be enormous in boosting up agricultural production[38]. Selective reports in which 2D analysis has proven to be of great use in plants are shown in Table 2. Here we consider in some detail the specific case of plant-abiotic stress interactions to further illustrate how proteomics research is proving to be of great help towards identifying novel stress-responsive genes and towards genetic engineering for increased-level stress tolerance. Production of abiotic stress-tolerant transgenic plants is critical for the much-needed future increase in crop production and yet this research is presently limited by dearth of information on genes related to stress tolerance[1,3,39–44]. The physiology

Table 2. Selected examples of the usage of two-dimensional protein gel electrophoresis technique in plant systems

Application	Reference
Identification of somatic embryogenesis-related proteins	76
Construction of wheat seed protein map	77
Scoring of polymorphism in *Saccharum* sp.	78
Association of protein amount polymorphism with performance of hybrids in maize	79
Identification of marker protein for distinguishing indica-japonica rice	80
Preparation of data-file of rice seed proteins	81
Tagging of plasma membrane proteins in *Arabidopsis*	82
Examining fate of the transgene product in *A. thaliana*	83
Identification of lumenal and peripheral proteins in pea thylakoids	84
Analysis of soybean peribacteroid membrane proteins	85

and biochemistry of plant abiotic stress responses is also poorly understood[43,45–47]. As a shot-gun approach to unravel biochemical and molecular changes elicited in stressed cells, electrophoresis of proteins isolated from non-induced (control) and induced (stressed) cells has been practised in a number of studies. The bulk of the early work on stress proteins was carried out employing 1-dimensional (1D) protein gel electrophoresis. Thus, the initial detection of 'pathogenesis-related proteins or PR proteins' and 'heat shock proteins or HSPs' in plants was made by this approach. However, more recently, 2D is being routinely employed for the objective of analysing stress proteins in plants. Costa *et al.*[48] employed 2D + micro-sequencing in order to identify the drought-responsive proteins that accumulate during the phase of water deprivation in *Pinus pinaster* seedlings. Of a total of 1000 protein spots resolved on the gel in this study, 38 responded to stress. When internal micro-sequences obtained for 11 proteins were analysed, 10 could be identified through sequence homology-based search. Importantly, the identified proteins were found to be associated with diverse processes such as photosynthesis, cell elongation, antioxidant metabolism and lignification.

The detailed characterization of stress proteins and their corresponding genes has proven to be of immense practical value. To appreciate this, let us first take here the example of flooding stress which is a major abiotic factor that affects several important crops[49,50]. The cellular damage in response to flooding stress is caused mainly due to anoxia as O_2 supply to submerged plant parts is drastically reduced[49]. Sachs *et al.*[51] analysed proteins induced in response to anaerobic stress in maize primary roots and reported that a small number of 'anaerobic polypeptides (ANPs)' accounts for more than 70% of total protein synthesis after 5 h of anaerobic stress. Further work established that ANPs are mostly constituted by the enzymes of glycolysis and ethanolic fermentation pathways[49], suggesting that anoxically-treated cells up-regulate ethanolic fermentation and glycolytic pathways as a strategy to survive under such stress conditions. These observations provided the basis for the recent work in which transgenic rice plants have been produced that over-expressed pyruvate decarboxylase (PDC), an enzyme that rate-limits the ethanolic fermentation process. Transgenic plants over-expressing PDC were found to possess relatively higher flooding tolerance[52]. There are several other examples of a similar nature, wherein useful research strategies have emerged based on proteome information. Moons *et al.*[53] have reported that 'late embryogenesis abundant (LEA)' proteins are present in higher amounts in the salt-tolerant rice cultivars compared to the sensitive cultivar. When the *hva1* gene of *Hordeum vulgare* (which encodes LEA protein) was transformed into rice cells, these plants maintained higher growth rates than the non-transformed plants (control) under water and salt stress conditions[54]. Earlier it was reported from our laboratory that 100 kDa HSPs are accumulated to a significant level in diverse plant species in response to high temperature stress[55–59]. When recently the *hsp100* gene was over-expressed in the transgenic *Arabidopsis* plants, it caused a significant increase in high temperature tolerance[60].

The future avenues for further increasing stress tolerance warrant that several stress tolerance-related genes must be pyramided[3]. The realization of this goal can only be achieved if major breakthroughs are made in further identification of the stress-related proteins and isolation and cloning of the requisite genes[3,44]. Genomics and proteomics research will be of great help in constantly expanding the information on newer stress-responsive genes and proteins. For instance, Moons *et al.*[61] have reported that complete submergence of rice seedlings for 60 h increased the accumulation of a 97 kDa protein in roots. When peptides generated by *in situ* tryptic digestion of this 2D protein spot were analysed, significant homology to plant pyruvate orthophosphate dikinase (PPDK) protein was revealed. This study thus associated PPDK protein to flooding stress response in rice. Recently, Chang *et al.*[62] analysed the patterns of protein synthesis during hypoxic and anoxia conditions in maize by employing 2D method. In this study, expression of as many as 262 individual proteins was shown to be altered with changes in O_2 tension regime. Further, of 48 protein spots analysed by MS, 46 were identifiable on the basis of database search. The identified proteins showed a wide range of functions. Thus, applications based on results of protein analysis are enormous in production of abiotic stress-tolerant transgenic plants.

Conclusions

There is a great deal of progress in cataloguing novel gene sequences accruing from the current international initiatives on the genome projects. Parallel efforts are being put to unveil the functionality of these genes by the approach of proteomics. High-throughput in proteomics is a must in this endeavour. The 2D technique is one of the basic inputs in proteomics research. The experimental methods dealing with extraction of proteins and their separation by 2D have been optimized to a great deal. While 2D works reasonably fast in terms of 'qualitative' performance, it is the 'quantitative' performance that counts in application of 2D in proteomics research. In the current method of analysis, each 2D spot is separately extracted, digested and analysed, which is a time-consuming process[19]. Additionally, 2D currently has an insufficient dynamic range for complete proteome analysis owing to its limited loading

capacity and the detection limits of staining. Further, specific classes of proteins are excluded or underrepresented in 2D gel patterns. These include very acidic or basic proteins, excessively large or small proteins and membrane proteins[19]. Clearly, the detection and quantification of low-abundance proteins such as transcription factors, protein kinases and other regulatory proteins is incompatible with the standard 2D + MS approach[14,30,63]. These arguments call for further technical developments in the 2D methods. It is suggested that development of 'protein chips', analogous to microarray chips for nucleic acids, could provide convenient high-throughput solution to proteome analysis[64]. However, as proteins are more complex and more diverse compared with nucleic acids, development of such chips for proteomics has proved difficult so far[11,64]. Finally, it is possible that other methods of protein analysis such as HPLC and capillary IEF (CIEF) may be combined with MS in the years to come, for meeting the deficiencies that are encountered in current 2D protocols[65].

Plant biotechnology research looks at proteomics research with great optimism. With the realization that complete genome sequence of *Arabidopsis* (a dicot system) has already been obtained and *O. sativa* (a monocot system) genome sequence will shortly become available, analysis of stress proteins and stress genes has acquired added significance. It should be possible to obtain pictures of stress-associated global changes in mRNA/protein alterations with these developments. Once that stage comes, efforts would be needed to distinguish the stress-responsive genes which impart stress tolerance and which do not impart stress tolerance through appropriate experimentation, such as by using transgenic technology. Availability of newer stress-tolerance genes may then lead to fresh avenues for production of genetically engineered high-level stress tolerant plants through the pyramiding approach. The present-day proteomics research promises a great deal for the agriculture of tomorrow.

1. Grover, A. and Minhas, D., *Proc. Indian Natl. Sci. Acad. Part B*, 2000, **66**, 13–32.
2. Gibson, S. and Somerville, C., *Trends Biotechnol.*, 1993, **11**, 306–313.
3. Grover, A., Sahi, C., Sanan, N. and Grover, A., *Plant Sci.*, 1999, **143**, 101–111.
4. Lockhart, D. J. and Winzeler, E. A., *Nature*, 2000, **405**, 827–835.
5. Maheshwari, S. C., Maheshwari, N. and Sopory, S. K., *Curr. Sci.*, 2001, **80**, 252–261.
6. Walbot, V., *Plant Physiol.*, 1999, **119**, 1151–1155.
7. Boucher, D. and Hofte, H., *Plant Physiol.*, 1998, **118**, 725–732.
8. Somerville, C. and Somerville, S., *Science*, 1999, **285**, 380–383.
9. Thillement, H. *et al.*, *Electrophoresis*, 1999, **20**, 2013–2026.
10. Smith, H. B., *Plant Cell*, 2000, **12**, 303–304.
11. Pandey, A. and Mann, M., *Nature*, 2000, **405**, 837–845.
12. Abbott, A., *Nature*, 1999, **402**, 715–720.
13. Li, J. and Assmann, S. M., *Plant Physiol.*, 2000, **123**, 807–809.
14. Gygi, S. P. and Abersold, R., in *Proteomics: A Trends Guide* (eds Blackstock, W. and Mann, M.), Elsevier Science, London, 2000, pp. 31–36.
15. Mann, M., *Natl. Biotechnol.*, 1999, **17**, 754–755.
16. O'Farrell, P. H., *J. Biol. Chem.*, 1975, **250**, 4007–4021.
17. Bauw, G. and Montagu, M. V., in *Differentially Expressed Genes in Plants: A Bench Manual* (eds Hansen, E. and Harper, G.), Taylor and Francis, London, 1997, pp. 105–117.
18. Packer, N. H. and Harrison, M. J., *Electrophoresis*, 1998, **19**, 1872–1882.
19. Gorg, A., in *Proteomics: A Trends Guide* (eds Blackstock, W. and Mann, M.), Elsevier Science, London, 2000, pp. 3–6.
20. Blackstock, W., in *Proteomics: A Trends Guide* (eds Blackstock, W. and Mann, M.), Elsevier Science, London, 2000, pp. 12–16.
21. Yates, J. R., *Trends Genet.*, 2000, **16**, 5–8.
22. Henzel, W. J., Billeci, T. M., Stuffs, J. T. and Wong, S. C., *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 5011–5015.
23. Binz, P. A. *et al.*, *Anal. Chem.*, 1999, **71**, 4981–4988.
24. Shevchenko, A. *et al.*, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 14440–14445.
25. Shevchenko, A., Loboda, A., Shevchenko, A., Ens, W. and Standing, K. G., *Anal. Chem.*, 2000, **72**, 2132–2141.
26. Crawford, M. E., Cusick, M. E. and Garrels, I. I., in *Proteomics: A Trends Guide* (eds Blackstock, W. and Mann, M.), Elsevier Science, London, 2000, pp. 17–21.
27. Hatzimanikatis, V., Choe, L. H. and Lee, K. H., *Biotechnol. Prog.*, 1999, **15**, 312–318.
28. Neubauer, G. *et al.*, *Nat. Genet.*, 1998, **20**, 46–50.
29. Rout, M. P., Aitchinson, J. D., Suprapto, A., Hjertaas, K., Zhao, Y. and Chait, B. T., *J. Cell. Biol.*, 2000, **148**, 635–651.
30. Vorm, O., King, A., Bennett, K. L., Leber, T. and Mann, M., in *Proteomics: A Trends Guide* (eds Blackstock, W. and Mann, M.), Elsevier Science, London, 2000, pp. 43–46.
31. Jensen, O. N., in *Proteomics: A Trends Guide* (eds Blackstock, W. and Mann, M.), Elsevier Science, London, 2000, pp. 36–42.
32. Tranter, D., *Drug Discov. Today,* 1999, **4**, 1–2.
33. Blackstock, W. P. and Weir, M. P., *Trends Biotechnol.*, 1999, **17**, 121–127.
34. Bhasin, V. K., *Nature*, 2000, **403**, 698.
35. Gutierrez, J. A., *Int. J. Parasitol.*, 2000, **30**, 247–252.
36. Ryu, D. D. and Nam, D. H., *Biotechnol. Prog.*, 2000, **16**, 2–16.
37. Saegusa, A., *Nature*, 1999, **401**, 313.
38. Dhand, R., *Nature*, 2000, **405**, 819.
39. Grover, A. *et al.*, *Curr. Sci.*, 1998, **75**, 689–696.
40. Dhaliwal, H. S., Kawai, M. and Uchimiya, H., *Plant Biotechnol.*, 1998, **15**, 1–10.
41. Khanna-Chopra, R. and Sinha, S. K., *Curr. Sci.*,1998, **74**, 25–34.
42. Bajaj, S., Targolli, J., Liu, L. F., Ho, T. H. D. and Wu, R., *Mol. Breed.*, 1999, **5**, 493–503.
43. Minhas, D. and Grover, A., *Proc. Indian Natl. Acad. Sci. Part B*, 1999, **65**, 33–50.
44. Grover, A. *et al.*, *Curr. Sci.*, 2001, **80**, 206–216.
45. Grover, A., Pareek, A. and Maheshwari, S. C., *Proc. Indian Natl. Sci. Acad. Part B*, 1993, **59**, 113–127.
46. Pareek, A., Singla, S. L. and Grover, A., in *Strategies for Improving Salt Tolerance in Higher Plants* (eds Jaiswal, P. K., Singh, R. P. and Gulati, A.), Oxford & IBH Publishing, New Delhi, 1997, pp. 365–391.
47. Singla, S. L., Pareek, A. and Grover, A., in *Plant Ecophysiology* (ed. Prasad, M. N. V.), John Wiley & Sons, 1997, pp. 101–127.
48. Costa, P., Bahrman, N., Frigerio, J-M, Kremer, A. and Plomion, C., *Plant Mol. Biol.*, 1998, **38**, 587–596.
49. Minhas, D. and Grover, A., *Plant Sci.*, 1999, **146**, 41–51.
50. Mohanty, H. K., Mallik, S. and Grover, A., *Curr. Sci.*, 2000, **78**, 132–137.
51. Sachs, M. M., Freeling, M. and Okimoto, R., *Cell*, 1980, **20**, 761–767.

52. Quimlo, C. A. *et al.*, *J. Plant Physiol.*, 2000, **156**, 516–521.
53. Moons, A., Bauw, G., Prinsen, E., Montagu, M. V. and Straeten, D. V., *Plant Physiol.*, 1995, **107**, 177–186.
54. Xu, D., Duan, X., Wang, B., Hong, B., David, T. H. and Wu, R., *Plant Physiol.*, 1996, **110**, 249–259.
55. Singla, S. L. and Grover, A., *Plant Mol. Biol.*, 1993, **22**, 1177–1180.
56. Singla, S. L. and Grover, A., *Plant Sci.*, 1994, **97**, 23–30.
57. Pareek, A., Singla, S. L. and Grover, A., *Plant Mol. Biol.*, 1995, **29**, 293–301.
58. Singla, S. L., Pareek, A. and Grover, A., *Plant Sci.*, 1997, **125**, 211–219.
59. Singla, S. L., Pareek, A. and Grover, A., *Plant Mol. Biol.*, 1998, **37**, 911–919.
60. Quiestch, C. Hong, S. W., Vierling, E. and Lindquist, S., *Plant Cell*, 2000, **12**, 479–492.
61. Moons, A., Valcke, R. and Montagu, M. V., *Plant. J.*, 1998, **15**, 89–98.
62. Chang, W. W. P., Huang, L., Shen, M., Webster, C., Burlingame, A. L. and Roberts, K. M., *Plant Physiol.*, 2000, **122**, 295–317.
63. Thomas, J. D. and Bogelan, R. A., in *Proteomics: A Trends Guide* (eds Blackstock, W. and Mann, M.), Elsevier Science, London, 2000, pp. 7–11.
64. Cahill, D. J., in *Proteomics: A Trends Guide* (eds Blackstock, W. and Mann, M.), Elsevier Science, London, 2000, pp. 47–51.
65. Washburn, M. P. and Yates, J. R., in *Proteomics: A Trends Guide* (Blackstock, W. and Mann, M.), Elsevier Science, London, 2000, pp. 27–30.
66. Tiselius, A., *Harvey Lect.*, 1937, **35**, 1939–1940.
67. Raymond, S. and Weintraub, L. S., *Science*, 1959, **130**, 711.
68. Svensson, H., *Acta Chem. Scand.*, 1961, **15**, 325–341.
69. Davis, B. J., *Acta. Sci.*, 1964, **121**, 404–427.
70. Shapiro, A. N., Vinulsa, E. and Maizel, J. V., *Biochem. Biophys. Res. Commun.*, 1967, **28**, 815–820.
71. Margolis, J. and Kenrick, J., *Nature*, 1967, **214**, 1334–1336.
72. Laemmli, U. K., *Nature*, 1970, **227**, 680–685.
73. Merril, C. R., Switzer, R. C. and van Keuren, M. L., *Proc. Natl. Acad. Sci. USA*, 1979, **76**, 4335–4339.
74. Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Gorg, A., Westermeier, R. and Postel, W., *J. Biochem. Biophys. Methods*, 1982, **6**, 317–319.
75. Hruschka, W. R., *Clin. Chem.*, 1984, **30**, 2037–2039.
76. Sung, Z. R. and Okimoto, R., *Proc. Natl. Acad. Sci. USA*, 1981, **78**, 3683–3687.
77. Anderson, N. G., Tollaksen, S. L., Pascoe, F. H. and Anderson, L., *Crop Sci.*, 1985, **25**, 667–674.
78. Ramagopal, S., *Theor. Appl. Genet.*, 1990, **79**, 297–304.
79. de Vienne, D., Maurice, A., Josse, J. M., Leornardi, A. and Damerwal, C., *Cell Mol. Biol.*, 1991, **40**, 29–39.
80. Saruyama, H. and Shinbashi, N., *Theor. Appl. Genet.*, 1992, **84**, 947–951.
81. Komatsu, S., Kajiwara, H. and Hirano, H., *Theor. Appl. Genet.*, 1993, **86**, 935–942.
82. Santoni, V. *et al.*, *Plant J.*, 1998, **16**, 633–641.
83. Getzoff, T. P., Zhu, Genhai, Bohnert, H. J. and Jensen, R. G., *Plant Physiol.*, 1998, **116**, 695–702.
84. Peltier, J. B., Frisco, G., Kalume, D. E., Roepstorff, P., Nilsson, F., Adamska, I. and Wijk, K. J., *Plant Cell*, 2000, **12**, 319–342.
85. Panter, S., Thomson, R., de Bruxelles, G., Laver, D., Trevabkis, B. and Udvardi, M., *Mol. Plant Microbe Interact.*, 2000, **13**, 325–333.
86. Suzuki, K., Itai, R., Suzuki, K., Nakanishi, H., Nishizawa, N-K., Yoshimura, E. and Mori, S., *Plant Physiol.*, 1998, **116**, 725–732.
87. Bradford, M. M., *Anal. Biochem.*, 1976, **72**, 248–254.
88. Krauss, M. R., Collins, P. J. and Blose, S. H., *Nature*, 1989, **337**, 669–670.
89. Hossain, A., Huq, E., Grover, A., Dennis, E., Peacock, W. J. and Hodges, T. K., *Plant Mol. Biol.*, 1996, **31**, 761–770.

ACKNOWLEDGEMENTS. We thank DBT and the National Agricultural Technology Project for financial support. H.D. thanks UGC for the research fellowship award.