REVIEW

Subcellular shotgun proteomics in plants: Looking beyond the usual suspects

Paul A. Haynes[1, 2] *and Thomas H. Roberts*[1]

[1] Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, NSW, Australia
[2] Australian Proteome Analysis Facility, Macquarie University, North Ryde, NSW, Australia

In this review we examine the current state of analytical methods used for shotgun proteomics experiments in plants. The rapid advances in this field in recent years are discussed, and contrasted with experiments performed using current widely used procedures. We also examine the use of subcellular fractionation approaches as they apply to plant proteomics, and discuss how appropriate sample preparation can produce a great increase in proteome coverage in subsequent analysis. We conclude that the conjunction of these two techniques represents a significant advance in plant proteomics, and the future of plant biology research will continue to be enriched by the ongoing development of proteomic analytical technology.

Received: March 2, 2007
Revised: April 27, 2007
Accepted: April 27, 2007

Keywords:
Liquid chromatography-tandem mass spectrometry / Rice / Sample preparation / Shotgun proteomics / Subcellular fractionation

1 How and why do we analyse proteins in plants?

While the blueprint for how to assemble a cell is contained in the genetic code, the bricks and mortar used in the building process, as well as the builders themselves and the maintenance workers, are predominantly proteins. These are the molecules in cells that are directly responsible for the maintenance of correct cellular function and, by implication, the continued viability of the organism. In recent years advances in technology have enabled the simultaneous study of a whole range of proteins expressed in a cell at a given time. This has spawned a new subdiscipline of protein biochemistry, known as proteomics, where the proteome is defined as the protein complement expressed by a genome of an organism, tissue or cell type at a given time following growth under specific conditions.

Many of the analytical tools used in proteomics still lag behind the analogous tools used in the analysis of both RNA and DNA. It is becoming routine to undertake large-scale identification and quantification of thousands of different RNA or DNA molecules in a single experiment using an array prepared from a single initial sample. This is done using techniques involving DNA and cDNA microarrays, differential display PCR and SAGE. It is not yet possible to perform the same type of experiments at the protein level, but recent advances have narrowed the gap to a certain extent.

Any discussion of proteomics in plants must start with consideration of two fundamental facts. First, plant genomes, like those of animals, contain tens of thousands of genes and hence are able to express vast numbers of proteins and isoforms [1–3]. Second, plants are incredibly complex living systems consisting of not only a series of interdependent major organs such as leaf, root, stem, flower and seed, but also a myriad range of both highly specialised tissues within each of those organs, and specialised organelles and compartments within individual cells. Thus, an ideal proteomics approach

Correspondence: Associate Professor Paul A. Haynes, Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, NSW 2109, Australia
E-mail: paul.haynes@cbms.mq.edu.au
Fax: +61-2-98506200

Abbreviations: GFP, green fluorescent protein; **iTRAQ,** isotope tags for relative and absolute quantitation; **Mudpit,** multidimensional protein identification technique; **RP,** reversed-phase; **RuBisCO,** ribulose bisphosphate carboxylase/oxygenase complex; **SCX,** strong cation exchange

would consist of a combination of the following features: high sensitivity, high throughput, analytical robustness, the ability to differentiate between differentially expressed proteins and the ability to analyse as many as possible of the proteins present in a given sample at a given time.

Taking these points into consideration, we will discuss in this review advances in two areas of proteomics – subcellular fractionation and shotgun proteomics. The conjunction of these two approaches promises to deliver tangible benefits for plant proteomics in the postgenomics era. The ability to delve deeper into the expressed proteome of plants, in a relatively high-throughput fashion, will not only reveal new biological insights, but will change the questions we can ask and the way we design experiments in the future. We will focus on the model plants *Arabidopsis thaliana* and rice, both of which have complete genome reference sequences, although the annotation of the *Arabidopsis* genome is currently at a more advanced stage than that of rice. A completely sequenced genome is essential for many of the proteomic analysis experiments we will discuss [1, 2].

Proteomics with other plant species remains more difficult than with the standard model eudicot and monocot, *Arabidopsis* and rice, but this situation is likely to improve in the near future. A complete genome sequence is now available for poplar (*Populus trichocarpa*) and the genome sequences of several other species, such as barrel medic and maize, are scheduled to be completed in the next few years (Table 1).

2 Subcellular fractionation in plants

The thousands of gene products in plants are all targeted efficiently to specific locations in the cell by elaborate targeting machinery. Families of closely related gene products are abundant in plants, with families of at least two members known to exist for at least 50% of genes. The proteins encoded by these genes often appear to be functionally redundant, but the crucial difference between them is often their ultimate cellular destination. Genetic manipulation in *Arabidopsis* has produced very large numbers of plant lines with gene knockouts and over expression phenotypes, and these can be used to explore the functions of individual genes. The problem with this approach is that there is often no information regarding exactly when and where to look for phenotypic changes among plant organs or tissues or within

Table 1. Current status of plant genome sequencing projects

Plant species	Genome size (Mb)	Products and public access to sequence	Project status	Agency
Barrel Medic (*Medicago truncatula*)	550	Gene-rich BAC sequence in GenBank http://www.medicago.org http://www.medicago.toulouse.inra.fr	In progress, completion anticipated in 2009	Medicago Genome Sequencing Project
False Brome (*Brachypodium distachyon*)	355	Whole genome shotgun sequence http://www.jgi.doe.gov/sequencing/why/CSP2007/brachypodium.html	In progress	DOE Joint Genome Institute
Grape (*Vitis vinifera*)	500	Whole genome shotgun sequence http://www.vitaceae.org/molecular.html	In progress	INRA (France) IASMA (Italy)
Maize (*Zea mays*)	2600	BAC and whole genome shotgun sequences deposited in GenBank http://www.gramene.org/Zea_mays/index.html	In progress, completion anticipatedin 2009	Maize Genome Sequencing Project
Poplar (*P. trichocarpa*)	480	Whole genome sequence http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html.	Complete	DOE Joint Genome Institute
Potato (*Solanum tuberosum)*	840	Gene-rich BAC sequence in GenBank http://www.potatogenome.net	In progress	The Potato Genome Sequencing
Sorghum (*Sorghum bicolour*)	736	Whole genome shotgun sequence http://www.jgi.doe.gov/sequencing/why/CSP2006/sorghum.html	In progress	DOE Joint Genome Institute
Soybean (*Glycine max*)	1115	BAC and whole genome shotgun sequences deposited in GenBank http://www.ncbi.nlm.nih.gov/Genbank	In progress	DOE Joint Genome Institute
Tomato (*Solanum lycopersicum*)	950	Gene-rich BAC sequence in GenBank http://www.sgn.cornell.edu/help/about/tomato_sequencing.pl	In progress	International Tomato Genome Sequencing Project

plant cells. If transgenic plant lines are to be accepted as commercial products, it will be essential to document and understand any changes in protein localisation or abundance that may have occurred as a result of transformation.

There is considerable interest in plant biology in families of channel-like and transport proteins in plants, as these represent the communication pathways between cells, compartments and the external environment. This network of genes has been studied extensively from the point of view of sequence homology and overall expression levels, yet little is known about specific localisation and functions for these proteins within distinct plant membrane systems. As an example, certain small organic compounds are required to cross various distinct cellular membranes in plant cells as they perform their signalling and metabolic functions. There are now more than 280 known organic solute cotransporters reported in *Arabidopsis*, sorted into 35 distinct families. Proteins with similar sequences, but expressed in discrete locations, are likely to be responsible for the spatially separated activities these transporters provide. In addition, a number of endogenous exported peptides are known to be involved in the control of plant development and differentiation, such as Cle and Cle-like elements [4], systemin [5] and propep1 [6].

Cellular fractionation studies have been undertaken for several decades in a wide variety of biological systems, and a wealth of biochemical knowledge has accumulated. In yeast and mammalian systems, a large number of gene products have been localised to specific subcellular organelles or compartments, usually based on antibody-tracking microscopy studies. This has been accelerated by the genomic revolution of recent years. Equivalent research on plants, despite the availability of the first two complete genome sequences, still suffers from a limited knowledge base of protein localisation. Comparative genomics is of some use in prediction of protein locations by comparison with yeast or mammalian systems, but the enormous diversity of plant genomic sequences means that many proteins are not amenable to cross-species comparisons. Rice is an illustrative example of this genomic diversity. Cultivated rice (*Oryza sativa*) has a genome of 390 Mb. There are, however, 24 known species of rice within the genus *Oryza*, 12 of which are included in the *Oryza* Map Alignment project (OMAP). None of these 12 species, other than *Oryza sativa*, has a complete genome sequence available, but a recent report described the construction of BAC libraries for each [7]. The estimated genome sizes of these rice species range from 340 Mb for *Oryza brachyantha* and 760 Mb for *Oryza rufipogon* to 1700 Mb for *Oryza minuta*.

2.1 Bioinformatics in assignment of subcellular locations

The simplest option for placement of plant genomic data in a relevant context in cell biology is the use of bioinformatic algorithms to predict protein localisation. A number of programs of this nature are widely available, such as TargetP [8], MitoPRED [9], Predotar [10], PSORT [11], WoLFPSORT [12], LumenP [13], SubLoc [14], signalP [15] and secretomeP [16]. Based on primary sequence similarites, proteins can be predicted to be localised to the nucleus, mitochondrion, plastid, peroxisome and ER. A significant limitation of many of these tools is their inability to predict localisation in other membrane compartments including vacuole tonoplasts and plasma membranes. Furthermore, comparing data outputs from analysis of large-scale sequencing efforts often shows very little consensus between programs, leaving only relatively small sets of proteins that are consistently predicted to be localised to a given organelle or compartment. An obvious reason for this discrepancy is the lack of high quality information used to train the algorithms in the first instance, which is reinforced by poor overlap between experimental datasets and consensus predictions – often around 50% or less. Despite these issues, bioinformatics prediction of protein subcellular localisation is usually the first choice method of analysis. It can frequently pinpoint a correct protein localisation, or at least provide enough information to direct further experiments.

2.2 Determining subcellular location using epitope tagging and microscopy

The alternative to predictive approaches is direct experimental analysis of subcellular compartments with the aim of locating proteins within cells. In the plants for which a complete genome sequence is available, it is now possible to pursue this on a large scale by proteomic analysis, rather than the traditional approach of one protein at a time. In the postgenomic era, systematic studies of protein localisation in a given species are generally performed using either a combination of epitope tagging and microscopy, or by subcellular proteomics. Knowing the genome sequence allows the *in vivo* synthesis of proteins with epitope tags or other reporter proteins attached that can be used for either affinity purification or for microscopy. Fusion proteins incorporating green fluorescent protein (GFP) are the best known and most widely used example of the utilisation of reporter proteins. High-throughput screening of GFP-labelled proteins on a genome-wide scale has been reported in yeast [17], but this report served to highlight some of the limitations of the technique. Expression of a tagged protein under a nonphysiological promoter runs the risk of the end product being artificially forced into a nonphysiological location [18]. Furthermore, addition of a substantial tag like GFP, or even a smaller epitope like a 6-His tag, can result in a nonphysiological location. This can be caused by several different mechanisms: the targeting information inherent in the native protein may be masked by the presence of additional sequence information [19]; the tag itself may have targeting capabilities which overcome the weak targeting signal of the native protein [20, 21]; the tag may contain a localisation signal that is not obvious in one species but becomes active when transferred to another, such as the inadvertent nuclear

localisation signal [22] in the TAP tagging vector which was discovered when it was transferred from yeast to plants [23]; or the tag may cause the end product to be processed differently, resulting in differences in both target location and protein function [24].

2.3 Subcellular proteomic analysis

A different experimental approach, which avoids the problems of fusion protein expression referred to above, is to directly identify the protein components of compartments purified by subcellular fractionation. This is now commonly known as 'subcellular proteomics', and is one of the fast growing areas of research in proteomics. Subcellular proteomics studies are typically undertaken by researchers with a history of working with a particular functionality located in a cellular compartment of interest. As an example, scientists studying photosynthetic processes would analyse the proteome of the purified chloroplasts in order to gain greater understanding of the key protein components and molecular mechanisms involved. Studies of this type have been reported for many different organelles, including in *Arabidopsis* for plastids [25, 26], mitochondria [27, 28], nuclei [29–31], peroxisomes [32], cell walls [33] and plasma membranes [34, 35]. The approach frequently used involves 2-D gel-based proteomics and has resulted in the identification and classification of a relatively small number of proteins, typically less than 100. This clearly represents only a small sample of the most abundant proteins that are actually expressed, and is most likely biased against an accurate representation of the number of membrane proteins present [36]. Despite recent advances in analysis of membrane proteins by using organic solvents [37] or novel detergents [38] in sample preparations for 2-D gels, significant limitations remain which have led to many researchers pursuing alternative methods.

Recent reports have included nongel-based MS/MS approaches to such analyses, also known as 'shotgun proteomics' methods. These identify a much larger number of proteins, such as the identification of 392 proteins in *Arabidopsis* chloroplast envelope membranes [39] and the identification of 268 proteins in *Arabidopsis* cell wall preparations [40], and contain a much higher proportion of very small proteins, very large proteins and integral membrane proteins [41]. The strength of these approaches is the inherent, and ever-increasing, sensitivity of the MS systems used, but that is also responsible for creating one of the problems. Contaminant proteins from other compartments are routinely being detected at low levels using sample preparation techniques that were previously considered adequate for producing 'pure' samples for analysis. There are, however, a small but significant number of proteins that are targeted to dual or multilocations [42]. This problem is exacerbated when contrasting results are reported by different researchers using varying techniques and tissues in analysis of what is nominally the same subcellular compartment. In the absence of any further information, it is impossible to know if the protein of interest should be considered as a 'contaminant' or assigned a new subcellular location at a lower expression level. The fact that sample purity is now the limiting factor in these experiments means that in-depth analysis of a single compartment is now likely to reveal the identity of all the major components expressed above a certain level, but some ambiguity is highly likely to remain regarding exact localisation of less abundant proteins.

2.4 Sample preparation from plant organelles

In order to attempt subcellular proteomic studies, it is necessary to have a high volume of starting material that is free from nonproteinaceous contaminants and amenable to fractionation approaches. Many researchers thus choose to proceed with cell culture systems rather than whole plants. One unintended side effect of this is often observed in the initial steps of sample preparation, where mechanical cell disruption usually breaks around half of the internal cellular compartments, making the soluble fraction a mixture of cytosol and the contents of many organelles. In etiolated plastids, as found in cultured cells, almost all the plastids are broken by mechanical disruption. This problem, however, can be alleviated by the use of an alternative approach that uses protoplast formation followed by gentle disruption and low-speed centrifugation [43]. One additional problem that often occurs during the isolation of subcellular compartments is proteolysis by endogenous proteases. This can give rise to artefacts in 2-D gel experiments, such as protein spots appearing at unexpected positions. While unintended proteolysis may not appear to be as much of a problem in preparing samples for shotgun proteomics experiments, prolonged incubation at elevated temperature can on occasion cause endogenous proteases to completely digest a significant proportion of the proteome.

A variety of methods are available for purification and isolation of subcellular compartments from *Arabidopsis* cells, many of which are based on differential centrifugation and gradient centrifugation on sucrose and percoll layers. These techniques have been used extensively in the preparation of mitochondria, with an initial step used to purify nuclei and subsequent high-speed centrifugation steps used to fractionate mitochondria, peroxisomes and plastids. Peroxisomes can be further purified using percoll cushion centrifugation [32], plasma membrane-enriched fractions can be prepared using two-phase partitioning [44], and other microsomal membrane fractions, such as tonoplast [45], golgi [46] and ER [47], can be prepared using linear sucrose and gradient sorbitol centrifugation. In several published reports, preparations of this type have been further purified using orthogonal approaches, such as free flow electrophoresis which separates proteins on the basis of charge rather than density [48, 49].

Regardless of the organelle or subcellular compartment being studied, a major advantage in preparing subcellular fractions for analysis by a shotgun proteomics approach is

that only 50–200 μg of total protein is required for a successful analysis, which is typically ten-fold less than is required for a more traditional 2-D gel-based study.

3 The current paradigm in plant proteomics: 2-D PAGE

2-D PAGE for the separation of complex protein mixtures was developed in 1970 [50] and is still one of the most widely used techniques in proteomics. Proteins are separated by their p*I* through a pH-gradient gel matrix (IPG) and then by molecular weight in the second dimension [51, 52]. The extra dimension of separation and the added area of the gel matrix allows for the clear separation of several thousand proteins under optimal conditions. 2-D PAGE attained its full potential when researchers used MS to identify the separated proteins, also known as protein spots (or just spots), within the gel [53]. Excision of a protein spot can be performed by hand with a razor blade, and the subsequent procedures for destaining [54], in-gel protein digestion with a protease such as trypsin [55], and the elution of peptides [56, 57] are relatively simple. The resulting peptide mixtures are usually analysed by MALDI-based TOF MS to determine PMF of the peptides from the spot [58], or ESI-based MS/MS [59]. The resulting spectra are searched against protein sequence database references and the output indicates the degree of confidence of assignment of the spectra to a peptide present in the database [60, 61]. Based upon the quality of the match, the identity of the protein can be deduced.

By analysing treated and control samples or samples from different time points with 2-D PAGE, a differential comparison of protein expression can be visualised. Quantitation can be added to the mix by using a variety of different *in vitro* labelling techniques (ICAT; isotope-coded affinity tag) [62], or using sensitive and differential gel stains (DIGE) [63, 64]. All of these steps, with the exception of running 2-D gels, are amenable to various types of automation, which allows hundreds of differentially resolved proteins to be quickly extracted, processed and identified. There is a substantial body of proteomics research reporting the protein accumulation, profiling patterns and identity of proteins from various tissue of *Arabidopsis* [39–41] and rice [42–44].

Depending on the tissue, treatment or plant species, anywhere from hundreds to thousands of proteins have been resolved and in general, many of these studies serve as catalogues for the accumulation of proteins in plant cells or organelles [26, 29, 34, 65, 66]. There are several examples of 2-D PAGE databases where the results are reported as digitised images of the gel with associated protein identification information (see for example, (http://gene64.dna.affrc.go.jp/RPD/main_en.html and http://semele.anu.edu.au/2d/2d.html). In the Australian rice anther protein database, there are more than 4000 protein spots within the p*I* range of 4–11 and molecular weight range of 6–122 kDa, corresponding to approximately 10% of the rice genome [67]. The Japanese database provides 21 reference maps of proteins from specific developmental tissues consisting of more than 13 000 spots [68].

As well as whole tissue studies there are also analyses of subcellular fractions specifically containing plasma membranes, chloroplasts, mitochondria or nuclei. These databases clearly illustrate that subcellular fractionation helps decrease the complexity of the whole tissue protein extracts. They also illustrate, however, some of the limitations of 2-D PAGE. In most samples, several thousand proteins can be displayed. In plants, especially green tissues, this number is often significantly lower. The highly abundant ribulose bisphosphate carboxylase/oxygenase complex (RuBisCO) proteins, which can be thought of as the plant analogue of albumin in mammalian serum and plasma preparations, overwhelm many of the low abundance proteins in photosynthetic tissue samples that would otherwise be clearly resolved [69, 70]. Albumin depletion strategies have been commercially available for several years, but the first commercially available RuBisCO depletion products were released only very recently and thus have not been widely adopted. Researchers have employed various strategies in attempts to overcome RuBisCO abundance, such as IPG strips with overlapping p*I* ranges to reduce sample complexity [71, 72]; however, this increases labour, cost and time, especially given the lack of automation in running 2-D gels and the bottleneck of gel image analysis.

The underrepresentation of basic, hydrophobic and membrane-spanning proteins remains an inherent limitation of 2-D PAGE [73]. This restricts the detection of nearly 30% of all cellular proteins, and is especially deleterious in plant cells which are heavily populated by various specialised membranous structures such as thylakoids and plasmodesmata [74]. There are, however, some advantages in using 2-D gels in terms of additional information gained from a single experiment, including protein molecular weight and p*I* values [52, 75], the detection of protein modifications and isoforms [76] and some indication of the level of proteolysis occurring, either in response to physiological stress or during sample preparation [77].

4 The concept of shotgun proteomics

Genomics has been legitimised by the entrenched technology platforms enabling high-throughput DNA sequencing and the biological insights provided by the related research. Minimal sets of genes that define life have been described [78, 79] and genomics has led to unparalleled advancements in phylogenetics [80–82], genetics [83] and breeding [84]. Genomics spawned proteomics as a word and as a discipline that is empowered by high-throughput protein separation techniques and protein identification by MS. By itself, genomics provides a list of genes in a cell, but imparts limited knowledge of gene expression or the functions of the expressed gene products. Transcriptomics provides valuable

information on gene expression; however, when genomics is coupled with proteomics, an understanding of protein accumulation and organisation can emerge. For the many organisms whose genomes have been sequenced, protein accumulation reference maps that provide a global snapshot of the protein constituents of cells in various states and conditions are being produced [85–88]. Plants are no exception [89, 90]. Plant biology has greatly benefited from genomics but the rate of proteomics discovery in plants has not kept up with genome sequencing, partly because of difficulties in obtaining proteins from the many developmentally different yet spatially convergent cell types in various plant tissues. Plant cells, compared to animal cells, are also compressed by vacuoles and rigid cell walls traversed by plasmodesmata and are crowded by many membranous plastids. These differences introduce new challenges for protein extraction and separation which are not frequently encountered by researchers performing analogous experiments in yeast, bacteria or cultured human and animal cells.

4.1 Multidimensional protein identification technique (Mudpit)

One of the major advances in proteomics in recent years, and the first of the group of technologies that have become known as shotgun proteomics approaches, is multidimensional protein identification technology, or Mudpit. This refers to a HPLC protein separation method that allows the high-throughput identification of peptides using online MS/MS [91, 92]. In contrast to 2-D gels, basic, hydrophobic and membrane-spanning proteins can all be identified with relative ease [91, 93]. Initial studies of this approach demonstrated that more membrane spanning proteins were identified than had been found in previous analyses of comparable samples using 2-D PAGE [94, 95]. In contrast to gel-based techniques where selected proteins are tryptically digested after separation, Mudpit analysis requires that all proteins in a sample are digested into peptides before the separation steps. The separated peptides are then sequentially eluted into the mass spectrometer and analysed in a high-throughput automatic fashion.

There are several disadvantages to the Mudpit approach that must also be considered. First, direct sample loading into a mass spectrometer means that detergents commonly used to isolate hydrophobic proteins must be avoided since these are readily ionised and can cause significant interference [96]. This can be avoided by the use of MS-compatible detergents in protein extraction protocols [97]. Second, the direct loading of biological samples onto biphasic columns can lead to column clogging or slow deterioration in column performance. Third, most mass spectrometers suffer to some degree from ion suppression effects which hinder the detection of low abundance ions coeluting with ions of much greater abundance. Fourth, data output comprises tens of thousands of tandem mass spectra, the analysis of which requires the use of software such as SEQUEST, MASCOT and Xtandem [60, 61, 98] or *de novo* sequencing programs such as PEAKS or MS-BLAST [99, 100] to deduce the amino acid sequence information from the spectra. High powered computing and vast data storage space are necessary in order to process the very large number of spectra in a reasonable time. Since the original mixture of proteins is digested into peptides, the protein information must be reconstructed. Reassembly of peptides into proteins in an unambiguous manner can be difficult, or in some cases not possible in the absence of further experimental information, as many of these proteins can share the same sequences. The only feasible way to generate a protein list is to follow rules of parsimony such that the smallest, most-logical non-redundant set of proteins is assembled from the peptides [101, 102]. Freely available programs such as DTASelect [103] and DBParser [104] organise the peptides into sets of candidate proteins, and ProteinProphet [105] can perform a similar organisation with the added benefit of incorporating a probability model that can be used to ascertain the likelihood that a protein assembly is correct.

In the original implementation of Mudpit, peptides are separated on two columns coupled directly to each other, and the outlet of the second column is directed to the inlet orifice of a mass spectrometer [106]. Columns are constructed from fused-silica capillaries pulled to a fine tip capable of spraying liquid drops of ionised peptides into the mass spectrometer source [107]. The columns are typically packed with strong cation exchange (SCX) resin and reversed-phase (RP) resins. Samples are loaded onto the mixed phase column offline with a pressure cell, before the assembly is connected to a HPLC pump. A solvent gradient is run to elute peptides that were not bound to the SCX material, and then a small salt step is applied for a short time to move the most weakly bound fraction of peptides from the SCX to the RP column. These peptides are then eluted from the RP column by another solvent gradient. This process is repeated sequentially, with a higher concentration of salt applied each time. Peptides are thus eluted from the two-phase column separated by two orthogonal properties, charge and hydrophobicity [91]. There are several variations on this protocol designed to minimise the problem of column clogging, which often occurs in the analysis of complex biological mixtures. These include split-phase columns where the sample is loaded onto the SCX phase only and washed extensively prior to connection to the RP column [108, 109], and columns configured with RP-SCX-RP that are also aimed at extending the peptide separation capacity [110, 111].

Mudpit as described above typically requires custom-made materials, including columns, column packing pressure cells and source platforms that interface the column to the MS inlet and the HPLC pump. This degree of specialisation and expertise has tended to keep this technology out of the hands of many plant researchers. Variations have been developed which include collecting offline fractions of the peptide mixture as they elute from a standard SCX column, and then analyzing fractions by RP [39, 112, 113]. Such an

approach can be implemented using an autosampler to make it a high-throughput technique. Offline Mudpit offers the advantage that a much larger amount of peptides can be loaded onto the larger SCX columns, with a corresponding increase in number of proteins identified, while online Mudpit offers the advantage of minimal sample handling, with the entire loaded sample directed into the MS. As a result, offline Mudpit is most applicable in cases where a large amount of sample is available, while online Mudpit is the best choice in cases where sample amount is strictly limited.

4.2 Quantitative Mudpit analysis

Quantification in Mudpit analysis is a difficult issue but much progress has been made in recent years. Mudpit analysis was previously thought to be essentially nonquantitative [114, 115], but in recent years numerous publications have demonstrated that Mudpit analysis can be used for studies involving differential comparative quantitation between two samples.

Relative quantitation in Mudpit analysis was initially demonstrated using $^{14}N/^{15}N$ isotopic labelling of sample pairs for comparison, which had the disadvantage of being mainly restricted to culturable microorganisms grown under defined nutrient regimes [116]. There have been several subsequent reports using this approach in higher organisms, including the use of isotopically labelled *E. coli* as a food source for *Caenorhabditis elegans* and *Drosophila melanogaster* [117], the labelling of cultured mammalian cell lines [118], and long-term metabolic labelling of adult rats with a diet enriched in ^{15}N [119]. An alternative approach which produces similar results is to use stable isotope labelling, usually ^{13}C, of amino acids in culture (SILAC), which also labels proteins *in vivo* [120]. One interesting method which has been applied to plants is known as subtle modification of isotope ratio proteomics (SMIRP), which offers a convenient approach to *in vivo* isotope labelling of plant proteins [121]. This technique uses partial rather than complete isotope incorporation to reveal structural changes in proteins, but has not yet been shown to be applicable to quantitative Mudpit analysis.

Quantification in Mudpit analysis can also be performed using *in vitro* labelling techniques such as isotope coupled affinity tags (ICAT) [62, 122] and isotope tags for relative and absolute quantitation (iTRAQ) [123, 124]. This enables the analysis of protein samples prepared from sources where quantitative *in vivo* labelling is not appropriate, including mature plant tissues. The iTRAQ approach has several disadvantages in that it requires expensive reagents and a mass spectrometer of relatively high resolution to distinguish the various diagnostic fragment ions. One major advantage of this approach, however, is that the multiplexed reagents offer the ability to quantitatively analyse four or eight samples, rather than just two. This enables, for example, the analysis of multiple time points, which was employed in a groundbreaking study on time-resolved MS of tyrosine phosphorylation sites in the human epidermal growth factor receptor signalling network [123]. Recent studies have shown that iTRAQ labelling is both compatible with offline Mudpit analysis [125, 126], and applicable to plant tissues, as demonstrated by a study of the defence response phosphoproteome in *Arabidopsis* leaf [127].

More recently, it has been shown that relative quantitation can be performed in the absence of any label, using metrics such as spectral counting or protein scores [116, 128]. This method has been applied to a number of mammalian systems, such as the characterisation of differential protein expression in synapses [129], and the identification of biomarkers for Down's syndrome in maternal serum [130] and preterm birth in human cervical–vaginal fluid [131]. Of particular relevance to the topic of this review, the label-free comparative Mudpit analysis method has also been used in a study of protein expression differences between normal and diseased mouse cardiac muscle tissue, where accurate quantitation was achieved by the use of prefractionation into subcellular components, coupled with multiple repeat Mudpit analyses [132].

Two caveats must be attached to all these quantitative approaches. First, multiple replicate analyses are required to produce statistically relevant results and avoid random sampling errors associated with Mudpit. This is most apparent in the analysis of more complex mixtures, where the number of peptides eluting at a given time may overwhelm the capacity of the MS instrument to fragment them all sequentially, but it can usually be alleviated by performing several replicates and considering only the consistent results [133, 134]. Second, there is still no definitive method for establishing the purity of a subcellular fraction or organelle preparation. A protein detected at very low levels in a particular subcellular fraction may be a contaminant from an adjoining compartment, or it may be expressed at low levels in the fraction of interest; unfortunately, there is no way to be certain which is correct. The use of enzyme assays to define purity of subcellular fractions is widespread, and remains the best option available [135]. This relies on the availability of an assay for an enzyme known to be expressed in only one location, which can cause similar problems to those described above, as enzyme activity detected in an adjoining compartment may be due to contamination or may be evidence of protein expression in an unexpected location. Again, there is no definitive way to determine which is correct.

4.3 SDS-PAGE – nano-LC-MS/MS

Another widely used shotgun proteomics approach that has become popular in recent years is the use of SDS-PAGE gels as a simple fractionation approach for proteins. This provides many of the advantages of Mudpit analysis, without the need for acquiring further equipment resources. The SDS-PAGE fractionation is followed by in-gel digestion of all the fractions, then RP nano-LC-MS/MS of peptides extracted

from each fraction. The most obvious benefit of this method is that SDS-PAGE gel resources are commonly available to most researchers, and are very easy to use. Another is that standard SDS-PAGE sample buffer solubilises proteins in complex mixtures much more effectively than the high-urea buffers commonly used in preparing protein samples for either 2-D gels or Mudpit. Boiling in SDS-PAGE sample buffer will extract measurable amounts of protein from just about any biological sample. After protein separation, the gel lane containing the proteins is excised and divided into slices (usually 32 or some other number readily compatible with robotic liquid handling procedures). The gel slice is destained if necessary, reduced, alkylated and digested with trypsin, and the resulting peptides are extracted. The peptides are separated on a standard reusable RP column and the eluent is then analysed by MS/MS.

Results obtained by separating proteins on 1D gels are similar to those obtained from Mudpit analysis. One study which compared a series of different shotgun proteomic approaches used for analysis of a yeast mixed organelle lysate found that the SDS-PAGE gel slice approach was significantly better than Mudpit analysis in terms of number of proteins identified [136]. The main disadvantages of this technique are that the offline protein fractionation and gel handling are relatively labour intensive, and the handling and digestion of multiple fractions introduces more opportunity for sample contamination.

Plant researchers have been among the most active in using the SDS-PAGE gel slice approach, and have had success with subcellular, organelle and membrane proteomics research [66, 137–139]. It has also been shown that many of the detergents that are commonly used to extract proteins from unique plant membrane structures are not compatible with Mudpit analysis. However, many of these are sufficiently removed by 1-D gel separation and therefore do not interfere with subsequent analysis. Interestingly, it has also been reported recently that the detergent dodecyl-β-maltoside, which is commonly used in extracting proteins from plant membranes, can be employed as part of a Mudpit sample preparation step if the extracted proteins are then precipitated using TCA [140]. Presumably, the acid efficiently cleaves the glycosidic linkage, and the reaction products no longer interfere in mass spectrometric analysis.

4.4 IEF – nano-LC-MS/MS

A shotgun proteomics approach that has been developed in recent years, but has yet to find much traction in plant research, is the use of IEF electrophoresis as a first fractionation step. This has been investigated at the protein level, where it was found to be inferior to the use of SDS-PAGE as an initial fractionation step, most likely due to the difference in solubilising strength of the sample preparations used. However, a number of researchers have found that IEF separation of peptides provides a powerful and versatile front end to shotgun proteomic analysis. The biological samples are first digested with trypsin as for a Mudpit experiment, then applied to an IPG IEF strip and separated under high voltage. Once the peptides have migrated to their p*I*, they are extracted into aqueous buffer, separated on a standard reusable RP column and the eluent is analysed by MS/MS.

There are several disadvantages to this approach. It is labour intensive and requires significant sample handling, again introducing opportunity for inadvertent contamination. This is exacerbated by the low percentage of acrylamide used in IEF strips rendering them much less compatible with robotic liquid handling systems than comparable preparations using SDS-PAGE gels. Also, biological samples must be solubilised and digested in buffers compatible with IEF, similar to a Mudpit experiment, which means detergent use must be minimised.

There are, however, several unique advantages to the use of IEF as a front end to a shotgun proteomics experiment. IEF introduces an additional dimension of information into the experimental data, as the p*I* of the peptide is strongly correlated with the final location on the IEF strip. This parameter can be used as an additional constraint in database searching programs, and has been shown to greatly decrease false positive assignment rates and thus produce more matches at higher confidence [141, 142]. In addition, the IEF strips have a very high protein loading capacity; it is not uncommon to read of 2-D gel studies starting with protein amounts in the range from 500 μg to 1 mg. It is axiomatic in shotgun proteomics that the more starting material used, the more protein identifications are generated at the end of the experiment.

The use of IEF separation of peptides in shotgun proteomics experiments has been demonstrated in, for example, *E. coli* [143], rat [143] and *Drosophila* [142]. A direct comparison with the SDS-PAGE gel slice shotgun approach used for analysis of *Drosophila* nuclei showed that the peptide IEF shotgun approach resulted in an increase in protein identifications of more than 40% [143]. Clearly, this is a powerful approach and, despite the limitations inherent in the sample preparation required, it is likely to find wide applicability in plant proteomics research in the near future.

5 Conclusions

Identification and quantification of large numbers of proteins in as little time as possible will become increasingly important in the future for plant proteomics. This will be facilitated by: advances in instrumentation, such as the development of higher resolution Orbitrap mass spectrometers [144] and hardware for novel peptide fragmentation chemistries (electron transfer dissociation [145]); and advances in computing resources including centralised online data analysis facilities such as the Australian Proteome Computational Facility (http://www.apcf.edu.au). We believe sample preparation will remain the most crucial step in plant proteomics. As we plunge headlong into the postgenomic

era, the search for new enabling technologies will become ever more intense. We have illustrated in this review that there are a number of such technologies and approaches currently available, and these are now ready to be used in various combinations in order to solve interesting biological problems. For plant proteomics in particular, there are now subcellular fractionation techniques available that can be readily used prior to analysis by assorted shotgun proteomics approaches. The application of this powerful combination will enable us to look closer than ever before at the inner workings of plant cells at the molecular level.

The authors wish to thank the Macquarie University Research Development Grant scheme for support, and P. H. would like to acknowledge the NSW Office of Science and Medical Research Biofirst Scheme for support, and thank Ron Thorn and David Clarke for continued support and encouragement. This work was supported in part by the Australian Government through the Major National Research Facilities program funding of the Australian Proteome Analysis Facility.

6 References

[1] AGI, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000, *408*, 796–815.

[2] Goff, S. A., Ricke, D., Lan, T. H., Presting, G. *et al.*, A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* 2002, *296*, 92–100.

[3] Yu, J., Hu, S., Wang, J., Wong, G. K. *et al.*, A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 2002, *296*, 79–92.

[4] Fiers, M., Ku, K. L., Liu, C. M., CLE peptide ligands and their roles in establishing meristems. *Curr. Opin. Plant. Biol.* 2007, *10*, 39–43.

[5] Lindsey, K., Casson, S., Chilley, P., Peptides: New signalling molecules in plants. *Trends Plant Sci.* 2002, *7*, 78–83.

[6] Huffaker, A., Pearce, G., Ryan, C. A., An endogenous peptide signal in Arabidopsis activates components of the innate immune response. *Proc. Natl. Acad. Sci. USA* 2006, *103*, 10098–10103.

[7] Ammiraju, J. S., Luo, M., Goicoechea, J. L., Wang, W. *et al.*, The Oryza bacterial artificial chromosome library resource: Construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus Oryza. *Genome Res.* 2006, *16*, 140–147.

[8] Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 2000, *300*, 1005–1016.

[9] Guda, C., Fahy, E., Subramaniam, S., MITOPRED: A genome-scale method for prediction of nucleus-encoded mitochondrial proteins. *Bioinformatics* 2004, *20*, 1785–1794.

[10] Small, I., Peeters, N., Legeai, F., Lurin, C., Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 2004, *4*, 1581–1590.

[11] Nakai, K., Horton, P., PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 1999, *24*, 34–36.

[12] Sprenger, J., Fink, J. L., Teasdale, R. D., Evaluation and comparison of mammalian subcellular localization prediction methods. *BMC Bioinformatics* 2006, *7*, S3.

[13] Westerlund, I., Von Heijne, G., Emanuelsson, O., LumenP – A neural network predictor for protein localization in the thylakoid lumen. *Protein Sci.* 2003, *12*, 2360–2366.

[14] Chen, H., Huang, N., Sun, Z., SubLoc: A server/client suite for protein subcellular location based on SOAP. *Bioinformatics* 2006, *22*, 376–377.

[15] Bendtsen, J. D., Nielsen, H., von Heijne, G., Brunak, S., Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 2004, *340*, 783–795.

[16] Tjalsma, H., Feature-based reappraisal of the *Bacillus subtilis* exoproteome. *Proteomics* 2007, *7*, 73–81.

[17] Proszynski, T. J., Klemm, R. W., Gravert, M., Hsu, P. P. *et al.*, A genome-wide visual screen reveals a role for sphingolipids and ergosterol in cell surface delivery in yeast. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 17981–17986.

[18] Piedras, P., Rivas, S., Droge, S., Hillmer, S., Jones, J. D., Functional, c-myc-tagged Cf-9 resistance gene products are plasma-membrane localized and glycosylated. *Plant J.* 2000, *21*, 529–536.

[19] Rial, D. V., Lombardo, V. A., Ceccarelli, E. A., Ottado, J., The import of ferredoxin-NADP+ reductase precursor into chloroplasts is modulated by the region between the transit peptide and the mature core of the protein. *Eur. J. Biochem.* 2002, *269*, 5431–5439.

[20] Zhou, J., Weiner, H., The N-terminal portion of mature aldehyde dehydrogenase affects protein folding and assembly. *Protein Sci.* 2001, *10*, 1490–1497.

[21] Chew, O., Whelan, J., Millar, A. H., Molecular definition of the ascorbate-glutathione cycle in Arabidopsis mitochondria reveals dual targeting of antioxidant defenses in plants. *J. Biol. Chem.* 2003, *278*, 46869–46877.

[22] Ma, C., Mitra, A., Intrinsic direct repeats generate consistent post-transcriptional gene silencing in tobacco. *Plant J.* 2002, *31*, 37–49.

[23] Rohila, J. S., Chen, M., Cerny, R., Fromm, M. E., Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. *Plant J.* 2004, *38*, 172–181.

[24] Haynes, P. A., Cross, G. A., Differential glycosylation of epitope-tagged glycoprotein Gp72 during the *Trypanosoma cruzi* life cycle. *Mol. Biochem. Parasitol.* 1996, *83*, 253–256.

[25] Schubert, M., Petersson, U. A., Haas, B. J., Funk, C. *et al.*, Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. *J. Biol. Chem.* 2002, *277*, 8354–8365.

[26] Peltier, J. B., Friso, G., Kalume, D. E., Roepstorff, P. *et al.*, Proteomics of the chloroplast: Systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell.* 2000, *12*, 319–341.

[27] Millar, A. H., Heazlewood, J. L., Kristensen, B. K., Braun, H. P., Moller, I. M., The plant mitochondrial proteome. *Trends Plant Sci.* 2005, *10*, 36–43.

[28] Millar, A. H., Liddell, A., Leaver, C. J., Isolation and subfractionation of mitochondria from plants. *Methods Cell Biol.* 2001, *65*, 53–74.

[29] Bae, M. S., Cho, E. J., Choi, E. Y., Park, O. K., Analysis of the Arabidopsis nuclear proteome and its response to cold stress. *Plant J.* 2003, *36*, 652–663.

[30] Calikowski, T. T., Meulia, T., Meier, I., A proteomic study of the arabidopsis nuclear matrix. *J. Cell Biochem.* 2003, *90*, 361–378.

[31] Khan, M. M., Komatsu, S., Rice proteomics: Recent developments and analysis of nuclear proteins. *Phytochemistry* 2004, *65*, 1671–1681.

[32] Fukao, Y., Hayashi, M., Nishimura, M., Proteomic analysis of leaf peroxisomal proteins in greening cotyledons of *Arabidopsis thaliana*. *Plant Cell Physiol.* 2002, *43*, 689–696.

[33] Chivasa, S., Ndimba, B. K., Simon, W. J., Robertson, D. *et al.*, Proteomic analysis of the *Arabidopsis thaliana* cell wall. *Electrophoresis* 2002, *23*, 1754–1765.

[34] Prime, T. A., Sherrier, D. J., Mahon, P., Packman, L. C., Dupree, P., A proteomic analysis of organelles from *Arabidopsis thaliana*. *Electrophoresis* 2000, *21*, 3488–3499.

[35] Tanaka, N., Fujita, M., Handa, H., Murayama, S. *et al.*, Proteomics of the rice cell: Systematic identification of the protein populations in subcellular compartments. *Mol. Genet. Genomics* 2004, *271*, 566–576.

[36] Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., Aebersold, R., Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. USA* 2000, *97*, 9390–9395.

[37] Phadke, N. D., Molloy, M. P., Steinhoff, S. A., Ulintz, P. J. *et al.*, Analysis of the outer membrane proteome of *Caulobacter crescentus* by two-dimensional electrophoresis and mass spectrometry. *Proteomics* 2001, *1*, 705–720.

[38] Luche, S., Santoni, V., Rabilloud, T., Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics* 2003, *3*, 249–253.

[39] Froehlich, J. E., Wilkerson, C. G., Ray, W. K., McAndrew, R. S. *et al.*, Proteomic study of the *Arabidopsis thaliana* chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. *J. Proteome Res.* 2003, *2*, 413–425.

[40] Bayer, E. M., Bottrill, A. R., Walshaw, J., Vigouroux, M. *et al.*, Arabidopsis cell wall proteome defined using multidimensional protein identification technology. *Proteomics* 2006, *6*, 301–311.

[41] Whitelegge, J. P., Plant proteomics: BLASTing out of a MudPIT. *Proc. Natl. Acad. Sci. USA* 2002, *99*, 11564–11566.

[42] Peeters, N., Small, I., Dual targeting to mitochondria and chloroplasts. *Biochim. Biophys. Acta* 2001, *1541*, 54–63.

[43] Madhusudhan, R., Ishikawa, T., Sawa, Y., Shigeoka, S., Shibata, H., Characterization of an ascorbate peroxidase in plastids of tobacco BY-2 cells. *Physiol. Plant* 2003, *117*, 550–557.

[44] Santoni, V., Rouquie, D., Doumas, P., Mansion, M. *et al.*, Use of a proteome strategy for tagging proteins present at the plasma membrane. *Plant J.* 1998, *16*, 633–641.

[45] Blumwald, E., Fortin, M. G., Rea, P. A., Verma, D. P., Poole, R. J., Presence of host-plasma membrane type H-ATPase in the membrane envelope enclosing the bacteroids in soybean root nodules. *Plant Physiol.* 1985, *78*, 665–672.

[46] Rochester, C. P., Kjellbom, P., Andersson, B., Larsson, C., Lipid composition of plasma membranes isolated from light-grown barley (*Hordeum vulgare*) leaves: Identification of cerebroside as a major component. *Arch. Biochem. Biophys.* 1987, *255*, 385–391.

[47] Berczi, A., Caubergs, R. J., Asard, H., Partial purification and characterization of an ascorbate-reducible b-type cytochrome from the plasma membrane of *Arabidopsis thaliana* leaves. *Protoplasma* 2003, *221*, 47–56.

[48] Zischka, H., Weber, G., Weber, P. J., Posch, A. *et al.*, Improved proteome analysis of *Saccharomyces cerevisiae* mitochondria by free-flow electrophoresis. *Proteomics* 2003, *3*, 906–916.

[49] Bardy, N., Carrasco, A., Galaud, J. P., Pont-Lezica, R., Canut, H., Free-flow electrophoresis for fractionation of *Arabidopsis thaliana* membranes. *Electrophoresis* 1998, *19*, 1145–1153.

[50] Kenrick, K. G., Margolis, J., IEF and gradient gel electrophoresis: A two-dimensional technique. *Anal. Biochem.* 1970, *33*, 204–207.

[51] Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E. *et al.*, Isoelectric focusing in immobilized pH gradients: Principle, methodology and some applications. *J. Biochem. Biophys. Methods* 1982, *6*, 317–339.

[52] Gorg, A., Obermaier, C., Boguth, G., Harder, A. *et al.*, The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 2000, *21*, 1037–1053.

[53] Yates, J. R., III, Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* 1998, *33*, 1–19.

[54] Gharahdaghi, F., Weinberg, C. R., Meagher, D. A., Imai, B. S., Mische, S. M., Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: A method for the removal of silver ions to enhance sensitivity. *Electrophoresis* 1999, *20*, 601–605.

[55] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* 1996, *68*, 850–858.

[56] Panigrahi, A. K., Allen, T. E., Stuart, K., Haynes, P. A., Gygi, S. P., Mass spectrometric analysis of the editosome and other multiprotein complexes in *Trypanosoma brucei*. *J. Am. Soc. Mass. Spectrom.* 2003, *14*, 728–735.

[57] Cooper, B., Eckert, D., Andon, N. L., Yates, J. R., III, Haynes, P. A., Investigative proteomics: Identification of an unknown plant virus from infected plants using mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2003, *14*, 736–741.

[58] Karas, M., Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* 1988, *60*, 2299–2301.

[59] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989, *246*, 64–71.

[60] Eng, J., McCormack, A. L., Yates, J. R., III, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Mass Spectrom.* 1994, *5*, 976–989.

[61] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551–3567.

[62] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. *et al.*, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 1999, *17*, 994–999.

[63] Komatsu, S., Zang, X., Tanaka, N., Comparison of two proteomics techniques used to identify proteins regulated by gibberellin in rice. *J. Proteome Res.* 2006, *5*, 270–276.

[64] Unlu, M., Morgan, M. E., Minden, J. S., Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis* 1997, *18*, 2071–2077.

[65] Kruft, V., Eubel, H., Jansch, L., Werhahn, W., Braun, H. P., Proteomic approach to identify novel mitochondrial proteins in Arabidopsis. *Plant Physiol.* 2001, *127*, 1694–1710.

[66] Millar, A. H., Heazlewood, J. L., Genomic and proteomic analysis of mitochondrial carrier proteins in Arabidopsis. *Plant Physiol.* 2003, *131*, 443–453.

[67] Imin, N., Kerim, T., Weinman, J. J., Rolfe, B. G., Characterisation of rice anther proteins expressed at the young microspore stage. *Proteomics* 2001, *1*, 1149–1161.

[68] Komatsu, S., Kojima, K., Suzuki, K., Ozaki, K., Higo, K., Rice Proteome Database based on two-dimensional polyacrylamide gel electrophoresis: Its status in 2003. *Nucleic Acids Res.* 2004, *32*, D388–D392.

[68] des Francs, C. C., Thiellement, H., de Vienne, D., Analysis of leaf proteins by two-dimensional gel electrophoresis: Protease action as exemplified by ribulose bisphosphate carboxylase/oxygenase degradation and procedure to avoid proteolysis during extraction. *Plant Physiol.* 1985, *78*, 178–182.

[70] Kim, S. T., Cho, K. S., Jang, Y. S., Kang, K. Y., Two-dimensional electrophoretic analysis of rice proteins by polyethylene glycol fractionation for protein arrays. *Electrophoresis* 2001, *22*, 2103–2109.

[71] Hoving, S., Voshol, H., van Oostrum, J., Towards high performance two-dimensional gel electrophoresis using ultrazoom gels. *Electrophoresis* 2000, *21*, 2617–2621.

[72] Wildgruber, R., Harder, A., Obermaier, C., Boguth, G. *et al.*, Towards higher resolution: Two-dimensional electrophoresis of *Saccharomyces cerevisiae* proteins using overlapping narrow immobilized pH gradients. *Electrophoresis* 2000, *21*, 2610–2616.

[73] Santoni, V., Molloy, M., Rabilloud, T., Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* 2000, *21*, 1054–1070.

[73] Douce, R., Joyard, J., Biochemistry and function of the plastid envelope. *Annu. Rev. Cell Biol.* 1990, *6*, 173–216.

[75] Haynes, P., Miller, I., Aebersold, R., Gemeiner, M. *et al.*, Proteins of rat serum: I. Establishing a reference two-dimensional electrophoresis map by immunodetection and microbore high performance liquid chromatography-electrospray mass spectrometry. *Electrophoresis* 1998, *19*, 1484–1492.

[76] Haynes, P. A., Phosphoglycosylation: A new structural class of glycosylation? *Glycobiology* 1998, *8*, 1–5.

[77] Imin, N., Kerim, T., Rolfe, B. G., Weinman, J. J., Effect of early cold stress on the maturation of rice anthers. *Proteomics* 2004, *4*, 1873–1882.

[78] Goffeau, A., Four years of post-genomic life with 6,000 yeast genes. *FEBS Lett.* 2000, *480*, 37–41.

[79] Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W. *et al.*, Life with 6000 genes. *Science* 1996, *274*, 546, 563–567.

[80] Chen, F. C., Li, W. H., Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* 2001, *68*, 444–456.

[81] Lutfiyya, L. L., Xu, N., DÓrdine, R. L., Morrell, J. A. *et al.*, Phylogenetic and expression analysis of sucrose phosphate synthase isozymes in plants. *J. Plant Physiol.*, in press, DOI: 10.1016/j.jplph.2006.04.014.

[82] Premzl, M., Gready, J. E., Jermiin, L. S., Simonic, T., Marshall Graves, J. A., Evolution of vertebrate genes related to prion and Shadoo proteins–Clues from comparative genomic analysis. *Mol. Biol. Evol.* 2004, *21*, 2210–2231.

[83] Dear, P. H., Bankier, A. T., Piper, M. B., A high-resolution metric HAPPY map of human chromosome 14. *Genomics* 1998, *48*, 232–241.

[84] Sweigard, J. A., Carroll, A. M., Farrall, L., Chumley, F. G., Valent, B., *Magnaporthe grisea* pathogenicity genes obtained through insertional mutagenesis. *Mol. Plant Microbe Interact.* 1998, *11*, 404–412.

[85] Alonso, J., Rodriguez, J. M., Baena-Lopez, L. A., Santaren, J. F., Characterization of the *Drosophila melanogaster* mitochondrial proteome. *J. Proteome Res.* 2005, *4*, 1636–1645.

[86] Hall, N., Karras, M., Raine, J. D., Carlton, J. M. *et al.*, A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. *Science* 2005, *307*, 82–86.

[87] Madi, A., Mikkat, S., Ringel, B., Ulbrich, M. *et al.*, Mass spectrometric proteome analysis for profiling temperature-dependent changes of protein expression in wild-type *Caenorhabditis elegans*. *Proteomics* 2003, *3*, 1526–1534.

[88] Nedelkov, D., Kiernan, U. A., Niederkofler, E. E., Tubbs, K. A., Nelson, R. W., Investigating diversity in human plasma proteins. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 10852–10857.

[89] Holmes-Davis, R., Tanaka, C. K., Vensel, W. H., Hurkman, W. J., McCormick, S., Proteome mapping of mature pollen of *Arabidopsis thaliana*. *Proteomics* 2005, *5*, 4864–4884.

[90] Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L. *et al.*, Proteomic survey of metabolic pathways in rice. *Proc. Natl. Acad. Sci. USA* 2002, *99*, 11969–11974.

[91] Washburn, M. P., Wolters, D., Yates, J. R., III, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 2001, *19*, 242–247.

[92] Wolters, D. A., Washburn, M. P., Yates, J. R., III, An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* 2001, *73*, 5683–5690.

[93] Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K. *et al.*, Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* 2006, *5*, 2339–2347.

[94] Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S., Garrels, J. I., A sampling of the yeast proteome. *Mol. Cell Biol.* 1999, *19*, 7357–7368.

[95] Perrot, M., Sagliocco, F., Mini, T., Monribot, C. *et al.*, Two-dimensional gel protein database of *Saccharomyces cerevisiae* (update 1999). *Electrophoresis* 1999, *20*, 2280–2298.

[96] Drexler, D., Barlow, D. J., Falk, P., Cantone, J. *et al.*, Development of an on-line automated sample clean-up method and liquid chromatography-tandem mass spectrometry analysis: Application in an in vitro proteolytic assay. *Anal. Bioanal. Chem.* 2006, *384*, 1145–1154.

[97] Cadene, M., Chait, B. T., A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins. *Anal. Chem.* 2000, *72*, 5655–5658.

[98] Craig, R., Beavis, R. C., TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* 2004, *20*, 1466–1467.

[99] Ma, B., Zhang, K., Hendrie, C., Liang, C., *et al.*, PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 2003, *17*, 2337–2342.

[100] Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A. *et al.*, Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal. Chem.* 2001, *73*, 1917–1926.

[101] Carr, S., Aebersold, R., Baldwin, M., Burlingame, A. *et al.*, The need for guidelines in publication of peptide and protein identification data: Working group on publication guidelines for peptide and protein identification data. *Mol. Cell Proteomics* 2004, *3*, 531–533.

[102] Nesvizhskii, A. I., Aebersold, R., Interpretation of shotgun proteomic data: The protein inference problem. *Mol. Cell Proteomics* 2005, *4*, 1419–1440.

[103] Tabb, D. L., McDonald, W. H., Yates, J. R., III, DTASelect and Contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* 2002, *1*, 21–26.

[104] Yang, X., Dondeti, V., Dezube, R., Maynard, D. M. *et al.*, DBParser: Web-based software for shotgun proteomic data analyses. *J. Proteome Res.* 2004, *3*, 1002–1008.

[105] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 2003, *75*, 4646–4658.

[106] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E. *et al.*, Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* 1999, *17*, 676–682.

[107] Gatlin, C. L., Kleemann, G. R., Hays, L. G., Link, A. J., Yates, J. R., III, Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry. *Anal. Biochem.* 1998, *263*, 93–101.

[108] Guelman, S., Suganuma, T., Florens, L., Swanson, S. K. *et al.*, Host cell factor and an uncharacterized SANT domain protein are stable components of ATAC, a novel dAda2A/dGcn5-containing histone acetyltransferase complex in Drosophila. *Mol. Cell Biol.* 2006, *26*, 871–882.

[109] Jin, J., Cai, Y., Yao, T., Gottschalk, A. J. *et al.*, A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex. *J. Biol. Chem.* 2005, *280*, 41207–41212.

[110] Wu, C. C., MacCoss, M. J., Howell, K. E., Yates, J. R., III, A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.* 2003, *21*, 532–538.

[111] Wu, C. C., MacCoss, M. J., Mardones, G., Finnigan, C. *et al.*, Organellar proteomics reveals golgi arginine dimethylation. *Mol. Biol. Cell* 2004, *15*, 2907–2919.

[112] Takatalo, M. S., Kouvonen, P., Corthals, G., Nyman, T. A., Ronnholm, R. H., Identification of new Golgi complex specific proteins by direct organelle proteomic analysis. *Proteomics* 2006, *6*, 3502–3508.

[113] Wienkoop, S., Glinski, M., Tanaka, N., Tolstikov, V. *et al.*, Linking protein fractionation with multidimensional monolithic reversed-phase peptide chromatography/mass spectrometry enhances protein identification from complex mixtures even in the presence of abundant proteins. *Rapid. Commun. Mass Spectrom.* 2004, *18*, 643–650.

[114] Hunter, T. C., Andon, N. L., Koller, A., Yates, J. R., III, Haynes, P. A., The functional proteomics toolbox: Methods and applications. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2002, *782*, 165–181.

[115] Yamaguchi, K., Beligni, M. V., Prieto, S., Haynes, P. A. *et al.*, Proteomic characterization of the *Chlamydomonas reinhardtii* chloroplast ribosome. Identification of proteins unique to th e70 S ribosome. *J. Biol. Chem.* 2003, *278*, 33774–33785.

[116] Zybailov, B., Coleman, M. K., Florens, L., Washburn, M. P., Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal. Chem.* 2005, *77*, 6218–6224.

[117] Krijgsveld, J., Ketting, R. F., Mahmoudi, T., Johansen, J. *et al.*, Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nat. Biotechnol.* 2003, *21*, 927–931.

[118] Gustavsson, N., Greber, B., Kreitler, T., Himmelbauer, H. *et al.*, A proteomic method for the analysis of changes in protein concentrations in response to systemic perturbations using metabolic incorporation of stable isotopes and mass spectrometry. *Proteomics* 2005, *5*, 3563–3570.

[119] Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E., Yates, J. R., III, Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal. Chem.* 2004, *76*, 4951–4959.

[120] de Godoy, L. M., Olsen, J. V., de Souza, G. A., Li, G. *et al.*, Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* 2006, *7*, R50.

[121] Whitelegge, J. P., Katz, J. E., Pihakari, K. A., Hale, R. *et al.*, Subtle modification of isotope ratio proteomics; an integrated strategy for expression proteomics. *Phytochemistry* 2004, *65*, 1507–1515.

[122] Chen, R., Pan, S., Yi, E. C., Donohoe, S. *et al.*, Quantitative proteomic profiling of pancreatic cancer juice. *Proteomics* 2006, *6*, 3871–3879.

[123] Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J. *et al.*, Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell Proteomics* 2005, *4*, 1240–1250.

[124] Hu, J., Qian, J., Borisov, O., Pan, S. *et al.*, Optimized proteomic analysis of a mouse model of cerebellar dysfunction using amine-specific isobaric tags. *Proteomics* 2006, *6*, 4321–4334.

[125] Keshamouni, V. G., Michailidis, G., Grasso, C. S., Anthwal, S. *et al.*, Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-

mesenchymal transition reveals a migratory/invasive phenotype. *J. Proteome Res.* 2006, *5*, 1143–1154.

[126] Melanson, J. E., Avery, S. L., Pinto, D. M., High-coverage quantitative proteomics using amine-specific isotopic labeling. *Proteomics* 2006, *6*, 4466–4474.

[127] Jones, A. M., Bennett, M. H., Mansfield, J. W., Grant, M., Analysis of the defence phosphoproteome of *Arabidopsis thaliana* using differential mass tagging. *Proteomics* 2006, *6*, 4155–4165.

[128] Old, W., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. *et al.*, Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell Proteomics* 2005, *4*, 1487–1502.

[129] Liao, L., Pilotte, J., Xu, T., Wong, C. C., *et al.*, BDNF induces widespread changes in synaptic protein content and up-regulates components of the translation machinery: An analysis using high-throughput proteomics. *J. Proteome Res.* 2007, *6*, 1059–1071.

[130] Nagalla, S. R., Canick, J. A., Jacob, T., Schneider, K. A. *et al.*, Proteomic analysis of maternal serum in down syndrome: Identification of novel protein biomarkers. *J. Proteome Res.* 2007, *6*, 1245–1257.

[131] Pereira, L., Reddy, A. P., Jacob, T., Thomas, A. *et al.*, Identification of novel protein biomarkers of preterm birth in human cervical-vaginal fluid. *J. Proteome Res.* 2007, *6*, 1269–1276.

[132] Kislinger, T., Gramolini, A. O., MacLennan, D. H., Emili, A., Multidimensional protein identification technology (MudPIT): Technical overview of a profiling method optimized for the comprehensive proteomic investigation of normal and diseased heart tissue. *J. Am. Soc. Mass Spectrom.* 2005, *16*, 1207–1220.

[133] Durr, E., Yu, J., Krasinska, K. M., Carver, L. A. *et al.*, Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. *Nat. Biotechnol.* 2004, *22*, 985–992.

[134] Liu, H., Sadygov, R. G., Yates, J. R., III, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* 2004, *76*, 41934201.

[135] Andon, N. L., Hollingworth, S., Koller, A., Greenland, A. J. *et al.*, Proteomic characterization of wheat amyloplasts using identification of proteins by tandem mass spectrometry. *Proteomics* 2002, *2*, 1156–1168.

[136] Breci, L., Hattrup, E., Keeler, M., Letarte, J. *et al.*, Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing. *Proteomics* 2005, *5*, 2018–2028.

[137] Ferro, M., Salvi, D., Riviere-Rolland, H., Vermat, T. *et al.*, Integral membrane proteins of the chloroplast envelope: Identification and subcellular localization of new transporters. *Proc. Natl. Acad. Sci. USA* 2002, *99*, 11487–11492.

[138] Maltman, D. J., Simon, W. J., Wheeler, C. H., Dunn, M. J. *et al.*, Proteomic analysis of the ER from developing and germinating seed of castor (*Ricinus communis*). *Electrophoresis* 2002, *23*, 626–639.

[139] Marmagne, A., Rouet, M. A., Ferro, M., Rolland, N. *et al.*, Identification of new intrinsic proteins in Arabidopsis plasma membrane proteome. *Mol. Cell Proteomics* 2004, *3*, 675–691.

[140] Lee, J., Cooper, B., Alternative workflows for plant proteomic analysis. *Mol. Biosyst.* 2006, *2*, 621–626.

[141] Cargile, B. J., Bundy, J. L., Freeman, T. W., Stephenson, J. L., Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification. *J. Proteome Res.* 2004, *3*, 112–119.

[142] Krijgsveld, J., Gauci, S., Dormeyer, W., Heck, A. J., In-gel isoelectric focusing of peptides as a tool for improved protein identification. *J. Proteome Res.* 2006, *5*, 1721–1730.

[143] Cargile, B. J., Talley, D. L., Stephenson, J. L., Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of p*I* predictability of peptides. *Electrophoresis* 2004, *25*, 936–945.

[144] Makarov, A., Denisov, E., Lange, O., Horning, S., Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J. Am. Soc. Mass Spectrom.* 2006, *17*, 977–982.

[145] Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 9528–9533.

