Available online at www.sciencedirect.com
SCIENCE DIRECT®
PHYTOCHEMISTRY
Phytochemistry 65 (2004) 1449–1485
www.elsevier.com/locate/phytochem

Review

Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments

Russell P. Newton [a,c,*], A. Gareth Brenton [b,c], Chris J. Smith [a,c], Edward Dudley [c]

[a] *Biochemistry Group, School of Biological Sciences, University of Wales Swansea, Wallace Building, Singleton Park, Swansea SA2 8PP, UK*
[b] *Mass Spectrometry Research Unit, Department of Chemistry, Grove Building, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, UK*
[c] *Biomolecular Analysis Mass Spectrometry (BAMS) Facility, Grove Building, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, UK*

Received 10 January 2004; received in revised form 6 April 2004
Available online 13 May 2004

Abstract

The genome of several species has now been elucidated; these genomes indicate the proteomic potential of the cell. While identification of genomes has been, and continues to be, a technically and intellectually demanding process, the identification of the proteome contains inherently greater difficulties. The proteome of each living cell is dynamic, altering in response to the individual cell's metabolic state and reception of intracellular and extracellular signal molecules, and many of the proteins which are expressed will be post-translationally altered. Thus if the purpose of the proteome analysis is to aid the understanding of protein function and interaction, then it is identification of the proteins in their final state which is required: for this mass spectrometric identification of individual proteins, indicating site and nature of modifications, is essential. Here we review the principles of the methodologies involved in such analyses, give some indication of current achievements in plant proteomics, and indicate imminent and prospective technical developments.

Contents

1. Introduction 1450
2. Principles 1451
3. Experimental strategy 1452
   3.1. Extraction 1452
   3.2. SDS–PAGE separation of proteins 1452
   3.3. Mass spectrometric identification of proteins 1453
4. Principles of mass spectrometry 1453
   4.1. Sensitivity 1454
   4.2. Mass spectrometry instrument principles 1454
   4.3. Ionization 1455
   4.4. Tandem mass spectrometry (MS/MS) 1456
5. Mass spectrometry in practice in proteomic studies 1457
   5.1. MALDI-TOF MS and peptide mass fingerprinting 1457

* Corresponding author. Tel.: +44-1792-295377; fax: +44-1792-295447.
E-mail address: r.p.newton@swansea.ac.uk (R.P. Newton).

0031-9422/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2004.04.015

5.2. Electrospray tandem mass spectrometry (ESI-MS/MS) sequencing of peptides for protein identification . . . 1460
5.3. MALDI versus electrospray . . . 1463
5.4. Alternatives to SDS–PAGE for protein analysis . . . 1464
5.5. The study of differential protein expression . . . 1466
5.6. The study of protein modifications . . . 1467
5.7. Investigation of protein–protein interactions . . . 1467

6. Applications in plant proteomic studies . . . 1468
6.1. General . . . 1468
6.2. Studies with organs and tissues . . . 1469
6.3. Studies with seeds . . . 1471
6.4. Organelles . . . 1472
6.5. Chloroplast . . . 1472
6.6. Mitochondrion . . . 1475
6.7. Other organelles . . . 1476
6.8. Plant proteins with specific properties . . . 1476

7. Future directions . . . 1477

References . . . 1479

1. Introduction

Before considering the role of mass spectrometry in studies of plant proteomics, we need first to consider what we mean by the term "proteomics" (Lefkovits, 2003). The proteome is the full complement of proteins expressed by a genome (Wasinger et al., 1995) at a specific point in time. If one accepts the claims of equipment purveyors, then study of the proteome will follow the path of the study of the genome, taking advantage now of high-throughput systems allowing automated proteome analyses via mass spectrometry. The reality is that true proteome analysis is still a distant objective; proteomics currently is the study of only some of the multiple proteins concomitantly extant, effectively the study of just selected parts of the proteome. Before deciding upon our techniques to determine which parts of a sample proteome we will select for, we must first appreciate why we face this limitation.

The genome of several species from the simplest, mycoplasm (Wasinger et al., 1995), to the most complex, man (McPherson et al., 2001), has now been elucidated. These genomes indicate the proteomic potential of the cell. While identification of genomes has been, and continues to be, a technically and intellectually demanding process, the identification of the proteome contains inherently greater difficulties. The first major difference between genome and proteome analysis is that the genome is static, while the proteome of each living cell is dynamic, altering in response to the individual cell's metabolic state and reception of intracellular and extracellular signal molecules. Thus while the genome enables a prediction of the proteome simply as the gene products, this cannot be described as the proteome, since we do not know which genes are expressed at any specific moment in time, and many of the proteins which are expressed will be post-translationally altered, by one or more of approximately 200 modifications (Black, 2003; Mann and Jensen, 2003; Jansen et al., 2002; Liebler, 2002). Thus if the purpose of the proteome analysis is to aid the understanding of protein function and interaction, then it is identification of the proteins in their final state which is required: for this mass spectrometric identification of individual proteins, indicating site and nature of modifications, is essential.

The second major difference concerns the relative amounts of the components within the genome and proteome (see dynamic range discussion in Section 2). As it is often the low-expression, rapid-turnover, proteins that are critically involved in dynamic cellular processes, for example signal-transduction mechanisms, this constitutes another severe difficulty. If the original identification of such a protein after its isolation could be viewed as "the needle in the haystack" experiment, then our task to monitor many such proteins in a proteomic study is by analogy finding several hundred needles in several thousand haystacks!

To depress the potential plant proteome analyst still further, the third problem is that the discussions above have been considered in the context of qualitative analyses, i.e. identifying which proteins are present. If we wish a true functional proteomics study, it will undoubtedly require quantitation: given that equimolar samples of different proteins produce diagnostic ions of

different peak heights in a mass spectrum, then we must again face an extreme technical and intellectual challenge.

Despite the problems alluded to above, interest in plant proteomics has increased rapidly in recent years. Areas of specific interest have included the study of protein variations in different plant organs (Mo et al., 2003; Bahrman and Petit, 1995), variations in response to physiological events (Gallardo et al., 2001), identification of unknown plant viruses due to their proteome (Cooper et al., 2003), and the identification of microtubule binding proteins in plants (Chan et al., 2003). A number of review articles have already been published concerning the study of the plant proteome (Kersten et al., 2002; Roberts, 2002; Thiellement et al., 2002; Rossignol, 2001; Whitelegge, 2003): the objective of this review is to detail the principles of the methodologies involved in such analyses, to give some indication of current achievements in plant proteomics, and to indicate imminent and prospective technical developments. This is not intended as a comprehensive review of proteomics, rather as a guide of choices to be made by the novice and a stimulant to review their current choices to practitioners of plant proteomics. For comprehensive reviews of the practical aspects of the techniques utilized in proteomics the reader is referred to Chapman (1996, 2000), Link (1999), Lilley et al. (2001), Walker (2002), Kannicht (2002), Westermeier and Naven (2002), Jebanathirajah et al. (2002), Aebersold (2003), Mann and Jensen (2003), Whitelegge (2003), Wu and Yates (2003), and to the web sites listed by Nakayama (2001).

2. Principles

The analysis of the protein complement of a sample, be it a whole organism, tissue, cell or organelle, typically now proceeds through the phases of extraction, prefractionation, separation, mass spectrometry and identification. While successive phases can be linked together providing an ''on-line'' process, involving for example hyphenated liquid chromatography and mass spectrometry, the overall process is not readily sequentially automated and usually requires disjointed manual steps. Although total automation is possible the protocols are sample specific and it is certainly currently difficult to conceive a universal 'straight through' protocol for protein analysis in proteomic studies. Thus typically protocols for a new sample type have to be developed by trial and error: in addition two other major problems have to be addressed.

For classical type protein and enzyme studies, the rationale employed has been to purify a single protein to the greatest practicable extent, ideally to homogeneity. As the sample material is an extremely complex mixture of tens of, or even, with the inclusion of post-translational modifications, perhaps hundreds of thousands of different proteins, together with a great(er) number of other diverse biomolecular species, including not only low molecular mass intermediates but other macromolecules in the form of nucleic acids, carbohydrates and lipids, such purification is no mean feat. It has been achieved by utilizing differences in solubility, shape, polarity, stability and size between the target protein and all contaminants. The same principle holds for the selection of the total protein complement with the fundamental difference that now all protein species are selected for rather than one. This is intrinsically more difficult, since the mixture of proteins have no common features in respect of the above five properties and many protein species may share common features with non-protein species rather than with other proteins. Fortunately however, in practice it is frequently not necessary to totally remove all the non-protein species as the mass spectrometer can partially act as a chromatograph.

The second problem concerns the dynamic range of the analysis. For example, if a species has a genome of 10^4 genes, then if all were simultaneously expressed we would be examining a potential proteome of 10^4 proteins before modification, or allowing for modifications ca. 10^5–10^6 proteins, although of course we would not expect all genes to be concomitantly expressed. Identifying such a number of different proteins is a daunting task: it is rendered far more difficult by virtue of the fact that while protein X is expressed as 10,000 copies, only 10 copies of Y may be expressed: in any non-specific analytical technique in which both X and Y are present in these proportions, the signal from Y will be undetectable, lost in background noise. Thus, since there is no PCR equivalent for replicating proteins, for the vast majority of proteomic analyses it is only the most abundant 10%, 15% or 20% of proteins which are monitored. To obviate this problem some form of prefractionation is required: for example in the routine analysis of serum proteins, the first stage is the removal and discarding of the major protein component, albumin (e.g. Mimetic® albumin removal, Prometic Life Sciences, Fullerton, CA, USA). The problem is illustrated in Fig. 1 (Newton, Geisbrecht and van Geyschem, unpublished observations). Fig. 1(a) is the 2D electrophoretogram of a brain tissue protein extract: Fig. 1(b) is the same extract separated by liquid chromatography with a series of fractions collected, the protein concentration of each normalized to the same level for each, the fractions then being pooled, concentrated, and rerun on 2D electrophoresis. Nearly three times as many proteins are now visible. The experiment highlights the problem of the dynamic range but does not provide a solution: having partially separated proteins by LC, 2D electrophoretic separation is now redundant for qualitative analyses, since the fractions obtained can be utilized for MS analysis with or without digestion, but 2D

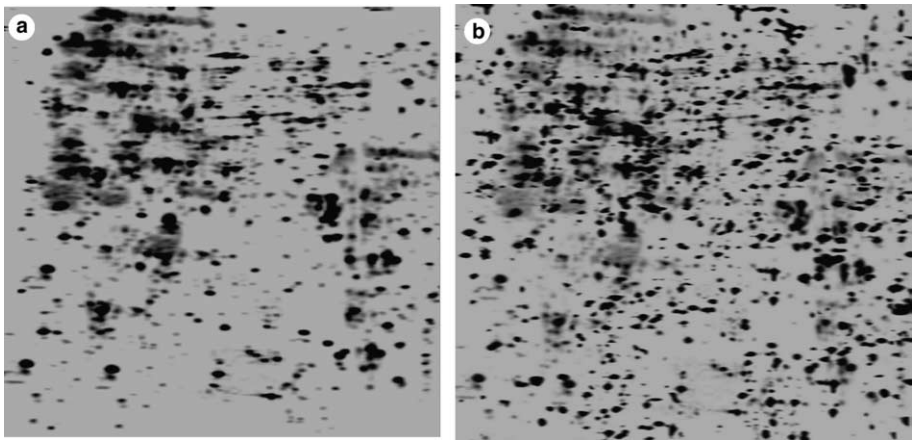


Fig. 1. Two-dimensional gel electrophoretograms of total protein extract (a) neat and (b) after chromatography and pooling of collected fractions following normalization of their protein concentrations.

electrophoresis still has an advantage that it provides the simplest means of quantitation, via image scanning/analysis software. However while 2D electrophoresis is still the cornerstone of most proteome analyses it is widely accepted that this role will rapidly be replaced by other separation methods.

3. Experimental strategy

3.1. Extraction

The extraction of the protein, unless the sample is an exudate of some type, will involve breakage of the cell and release of the contents into a soluble medium. The cleavage of the cell membrane typically involves physical disruption by mechanical means, sonication, or chemical or enzymic lysis. Classical purification of individual proteins prescribes distributing the target analyte protein into a buffered mixture as close as possible in composition to the original intracellular medium, for example in respect of pH and ionic strength, with the inclusion of additional components that will help solubilize the protein, protect it from hydrolysis or oxidation, and also remove non-proteinaceous constituents of the aqueous extract. Thus lipids can be removed by detergents, which also aid solubilization of membrane proteins, nucleic acids by selective pH precipitation, carbohydrates by specific enzymes, while proteins are protected by antioxidants such as dithiothreitol and mercaptoethanol and by protease inhibitors. There are many such recipes, both rationale and detailed methodology being reviewed by Deutscher (1990) and Link (1995).

The problem facing the proteomicist is that many of the succeeding separation and analytical steps may be intolerant of these additives. For example inorganic salts will interfere in electrospray mass spectrometry, detergents in chromatographic and electrophoretic separations and in MALDI mass spectrometry, while protease inhibitor cocktails may interfere in the digestion of the analyte proteins by trypsin or other exogenous proteins. It is thus essential to design extraction strategies with full knowledge of the nature and sensitivities of further processing and analytical stages.

3.2. SDS–PAGE separation of proteins

The majority of research into the plant proteome has utilised two-dimensional sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) for the protein separation stage. While SDS–PAGE still arguably has the greatest resolving power for the separation of complex mixtures of proteins (Thiellement et al., 2002), it suffers from disadvantages, particularly of dynamic range, discussed above in this review. During a SDS–PAGE separation the proteins are initially separated by isoelectric focusing (IEF) in which they migrate along an IEF strip which has a pH gradient between a cathode and an anode, the migration of each protein ending when it reaches its isoelectric point in the gradient. This strip is then applied to a polyacrylamide gel in which the second dimension of the separation occurs. Prior to this separation the proteins are unfolded and "coated" by SDS molecules, the effect of which is to mask the protein's original charge and give the protein a net negative charge. The negatively charged proteins are then attracted to a positive charge at the gel end. The mobility of the proteins is related to the unfolded chain length with smaller proteins moving faster and further through the gels. Once separated the proteins are generally visualised using a staining procedure such as coomassie blue or silver stain. The gels pieces that contain the proteins of interest are then excised and washed in order to remove the staining material used. A protease, usually trypsin, is then added in suitable con-

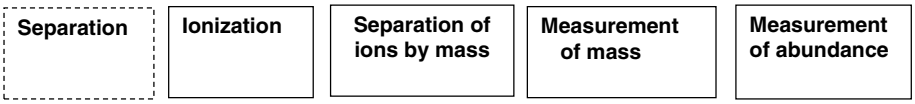


Fig. 2. Illustration of the four main functions of a mass spectrometer.

ditions to allow the digestion of the protein into peptides which are then washed out of the gel, collected and dried prior to further analysis, although undigested proteins can also be analysed by MALDI MS (see below). For a comprehensive appraisal of 2D electrophoresis the reader is directed to Gorg et al. (2000).

3.3. Mass spectrometric identification of proteins

In most cases one or both of two forms of mass spectrometry, which utilize different methods of sample ionisation, are used for protein identification. The first is matrix-assisted laser desorption ionisation (MALDI)-time-of-flight (TOF) mass spectrometry, used to perform peptide and protein mass fingerprinting; the second is electrospray (ESI) tandem mass spectrometry (MS/MS), usually hyphenated to high performance liquid chromatography (HPLC) separation, used to perform peptide sequence elucidation and identification of the corresponding protein.

4. Principles of mass spectrometry

Mass spectrometry has become an essential analytical tool in modern biological sciences, due not only to its outstanding sensitivity but also to the total information content delivered by the technique. Mass spectrometry has made rapid progress as an analytical technique, particularly over the last decade, with many new types of instrumentation being introduced. Constant refinements to sensitivity, selectivity and mass measurement accuracy have transformed mass spectrometry, leading to its wide scale uptake by the life sciences community with many reviews being published on the method of mass spectrometry and its applications to biological sciences. Recent reviews have surveyed the type of instrumentation that is currently dedicated to proteomics research (Binz et al., 2003) and the wide range of proteomes investigated (Figeys, 2003).

The principles of mass spectrometry can be envisaged by the four functions of the mass spectrometer outlined in Fig. 2. Ionization is fundamental as the physics of mass spectrometry relies upon the molecule of interest being charged, resulting in the formation of either a positive ion or negative ion. Whilst this is a strength of the technique it is also a weakness, as the connection between "the ion" and the neutral moiety it must represent may be removed as ionization is a suprathermal process imparting internal energy to the ion which can lead to rearrangement and even fragmentation. Upon ionization a molecular ion species (generally an adduct ion) is formed and, depending on the ionization method, fragment ions may also be created. A range of ionization techniques is available, listed in Table 1, together with their principal attributes and suitability for proteomic study. Next, these ion species are separated according to their mass-to-charge (m/z) ratio, and the masses assigned from the measurement of some physical parameter. Usually the mass is measured to better than 0.4 u so that a nominal mass can be assigned and in certain cases the mass is measured to much higher accuracy so that an accurate mass measurement is made. The latter is usually necessary when the elemental composition is required and can be applied to molecules of MW $<$800 u (above this mass there are usually too many combinations of the elements, for any given mass, to assign a unique elemental formula). Finally, the measurement of

Table 1
Ionization techniques in common usage in mass spectrometry and some of their properties in connection with proteomics

Ionization technique	Suitable for proteomics	Typical ions	Upper limit of mass range[a]	Usage
1. Electrospray/microspray/nanospray	Yes	Multiply protonated $[M+nH]^{n+}$	10^5–10^6u	Wide
2. Matrix-assisted laser desorption/ ionization (MALDI)	Yes	Singly protonated, $[M+H]^+$	10^5–10^6 u	Wide
3. Atmospheric pressure chemical ionization (APCI)	No	$[M+H]^+$	$<$2000 u	Small chemicals; excellent for quantitation
4. Atmospheric pressure photoionization (APPI)	No	$M^{+\cdot}$	$<$2000 u	Extends the compound range of APCI
5. Electron impact ionization	No	$M^{+\cdot}$	$<$3000 u	Synthetic/organic chemistry
6. Chemical ionization	No	$M+NH_4^+$	$<$2000 u	Synthetic/organic chemistry

[a] The mass range for all techniques is essentially from 0 Da to the mass range upper limit, which is dependent upon sample and instrument type.

ion abundance, based on peak height or peak area, is made leading to a semi-quantitative or quantitative answer.

4.1. Sensitivity

Whilst mass spectrometry is not the most sensitive of all analytical techniques the continual improvements being made to it have great potential for the future. In principle the ultimate sensitivity would be one ion per molecule; current cited limits of detection are in the low zepto mol range (Shen et al., 2004; and publications cited therein) corresponding to one ion per 6000 sample molecules, refer to Table 2. In general sensitivities are typically in the pmol to fmol range. The issue of sensitivity (limit of detection) is fraught with problems, often resulting from extreme claims in the literature involving simple and ideal sample preparations, rather than real world samples requiring purification and separation. Nevertheless the future promise of mass spectrometry is huge, particularly with the continuing improvements to sensitivity and selectivity current and expected. The overall efficiency (ε) of detecting a sample molecule, as an ion, by mass spectrometry is given by

$$\varepsilon = S_{\text{eff}} I_{\text{eff}} T_{\text{int}} T_{\text{ms}} D_{\text{eff}}, \quad (1)$$

where S_{eff} is the effective extraction and separation efficiency, i.e. the fraction of sample material presented to the mass ionization source; I_{eff} is the ionization efficiency; T_{int} is the transmission efficiency of the mass spectrometer interface (transfer from ion source to mass analyzer); T_{ms} is the transmission efficiency of the mass spectrometer analyzer; and D_{eff} is the detection efficiency of the ion detector system.

Typical values for these parameters are given in Table 3. In practice samples are contained in a matrices giving rise to chemical noise and the detection limits reduce under such circumstances by as much as 1–3 orders of magnitude depending upon the specific application.

4.2. Mass spectrometry instrument principles

There are five principal types of mass analyzer (listed in Table 4). The magnetic sector mass analyzer is included here on a historical basis; while not suitable for proteome study it is still a very good choice of analyzer for single sample analysis (MW <3000 u). The remaining four analyzers are: quadrupole; ion trap (IT) [either quadrupole ion trap (QIT) or linear quadrupole ion trap (LQIT)]; time-of-flight (TOF); and Fourier-transform ion cyclotron resonance (FTICR). They are used for mass analysis and their attributes at the listed in Table 5. The quadrupole and ion trap are very robust, low cost, but only offer low mass resolution. TOF and FTICR provide accurate mass and higher mass range, with the

Table 2
Molar quantities commonly used in mass spectrometry referring to the amount of substance and the number of molecules corresponding to each

No. of moles	Micro – 10^{-6}	Nano – 10^{-9}	Pico – 10^{-12}	Femto – 10^{-15}	Atto – 10^{-18}	Zepto – 10^{-21}	Yocto – 10^{-24}
No. of molecules	6×10^{17}	6×10^{14}	6×10^{11}	6×10^{8}	6×10^{5}	602	0.6

The normal amounts of sample used are shown in the hatched area.

Table 3
Typical values for the instrument parameters affecting the overall efficiency of detecting a sample by mass spectrometry

Parameter	S_{eff}	I_{eff}	T_{int}	T_{ms}	D_{eff}
Numerical value	$\leqslant 1$	~ 0.005[a]–0.00001	$\leqslant 0.1$	~ 0.1–0.5[b] ~ 0.001[c]	~ 1

[a] It is difficult to distinguish the combined effects of ionization and transmission efficiency of the ion optical interface in modern instrumentation. The product $I_{\text{eff}} T_{\text{int}}$ is typically no more than 0.001. These parameters are largely unreported in the literature and vary depending on the particular ionization method an interface type employed.

[b] Non-scanning instrument such as TOF have very high transmission >0.1.

[c] Scanning instruments – quadrupole and ion trap.

Table 4
The main mass analyzers used in mass spectrometry and features of note

Mass analyzer type	Suitable for proteomics	Robustness	Cost	Complexity	Speed
1. Magnetic sector	No	–	–	–	–
2. Quadrupole	Yes	Very	Low	Low	Fast
3. Ion trap (IT)	Yes	Very	Medium	Low-medium	Fast
4. Time-of-flight (TOF)	Yes	Moderate	Medium-high	Low-medium	Very fast
5. Fourier-transform ion cyclotron resonance (FTICR)	Yes	Moderate	High	High	Moderate

Table 5
Attributes and typical range of specifications of commercial mass spectrometer instruments

Mass analyzer type	Measures	Mass-to-charge (m/z) range	Mass resolution at 1000 Th[a,b]	Mass measurement accuracy (ppm)[c]	Dynamic range[d] (upper limit)	Sensitivity (mol)
1. Magnetic sector	Momentum-to-charge	10^4	10^5	2–5	10^7	–
2. Quadrupole	Path stability	10^4	10^3	~100	10^5	$\sim10^{-15}$
3. Ion trap (IT)	Frequency	10^3–10^4	10^3–10^4	~100	10^5	$\sim10^{-15}$
4. Time-of-flight (TOF)	Flight time	10^6	10^3–10^4	5–10	10^2–10^5	10^{-15}–10^{-13}
5. Fourier-transform ion cyclotron resonance (FTICR)	Cyclotron frequency	10^5	10^6	1–2	10^4	10^{-19}–10^{-15}

Stability of mass measurement accuracy can be poor with a TOF and some are very susceptible to temperature fluctuations. Internal calibration is essential.

[a] Th – Thomson the unit for mass-to-charge ratio.

[b] Mass resolution is usually defined at half peak height except for the magnetic sector instrument where the definition is 10% valley.

[c] ppm – Parts per million resolution.

[d] Typical best values; the wide range of values for TOF depends on the method for ion signal detection. Time-to-digital (TDC) detectors have lower dynamic range then analogue-to-digital based detectors. Hybrid TOF detectors can achieve up to 7 orders of magnitude dynamic range. For a typical acquisition of 1 s over a chromatographic peak the dynamic range would be 10^2–10^3, for the three most important techniques of IT, TOF and FTICR.

TOF having theoretically unlimited mass range. Within the listed attributes of each analyzer (Table 5) the typical range of each is represented. However some of these specifications may not be met easily under proteomic measuring conditions, due to restrictions on time of analysis for a mass spectrum. For example, a chromatographic peak of width 2 s may require up to 10 measurements for analysis, thus each mass spectrum should be no more than 0.2 s. Another restriction will be that some designs of TOF analyzer only achieve a restricted dynamic range (2–3 orders of magnitude), when running under fast separation conditions.

4.3. Ionization

Electrospray ionization (ESI) together with its low-flow variants, such as nanoflow-ESI, continues to dominate as one of the two preferred ionisation methods for proteomics (Griffiths et al., 2001). One issue with these ionization techniques is that the "molecular ion" (radical cation) is hardly ever seen, rather the ionized molecule is represented by an adduct ion, usually a singly or multiply protonated species. Fig. 3 illustrates some of the more common adduct ions that can be formed, and in this example a wide range of common adduct species are envisaged. Thus the molecular ion species is "diluted" amongst a number of adduct species and has the effect of lowering sensitivity. Knowledge and control of adducts formed during a proteomics experiment is therefore important.

Another physical effect comes from the natural isotopic distribution of a species. For example Fig. 4 shows the isotopic distribution of (a) low, (b) medium and (c) high mass ions. As molecular mass increases the monoisotopic ion becomes a smaller and smaller component of the isotopic envelope representing the molecular species; again "diluting" the molecular species over a range of masses. In addition, each nominal mass represented in the spectra (in Fig. 4(c)) can arise from a number of different combinations of elements, which become increasingly numerous with increasing mass and therefore has an important effect on the individual peak shape, at each nominal mass, and also on any associated mass measurement performed. A method has been described (Tang et al., 2002) to improve the apparent mass resolution, of low resolution mass spectra of biopolymers, and improve the identification of isotopic profiles, by using ^{12}C and ^{14}N enriched samples.

Electrospray ionization is also characterized by a distribution of charged states (roughly 1 charge per 1000 mass units, an illustration of the mass spectrum of a multiply charged peptide of molecular weight ~17,000 is shown in Fig. 5. This again leads to a "dilution" of the molecular ion species and a reduction in sensitivity if detection is based on one mass only. For applications where the charged state distribution can be connected to a single protein deconvolution procedures can be applied leading to molecular weight assignment (Ferrige et al., 1992). Charge state reduction (Scalf et al., 2000) has been developed where a distribution of charge state, typically <10 charges, can be condensed to a singly and doubly charged species. The method involves spraying the multiply charged ions into a specially designed charge reduction chamber placed at the entrance of the mass spectrometer.

It is also well established that the sensitivity of electrospray increases as the flow rate of the spray reduces: miniaturization of spray devices (Shen et al., 2004; Valaskovic and Kelleher, 2002) has been described and their improvement on signal intensity in LC/MS applications has shown sub attomol detection limits can be achieved.

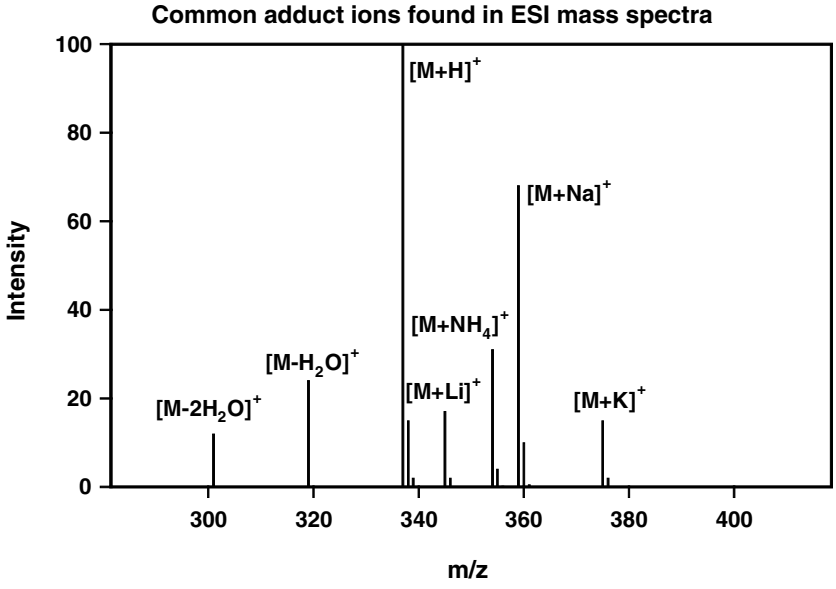


Fig. 3. Illustration of some of the common adduct ions observed in electrospray ionization for a compound of MW = 336 u. Common adducts in positive ion mode are: H (+1); Li (+7); NH_4 (+18); Na (+23); K (+39). Other adducts are the proton bound dimer $[2M + H]^+$ and multiply protonated species $[M + nH]^{n+}$. Common losses are one or two water molecules. In negative ion mode common ions are: $[M - H]^-$; $[M - Cl]^-$; $[M - 2H + Na]^-$; $[M - 2H + K]^-$; $[M + CH_3COO]^-$.

The second major ionization technique employed in proteomics is MALDI, which over the last decade it has been refined in a number of ways especially for biological applications. Several inter comparative studies have contrasted LC-MALDI-MS and LC/ESI-MS (for example, see Person et al., 2003) with the protein coverage of each technique being compared; very broadly the ESI/nanospray and MALDI have similar coverages.

4.4. Tandem mass spectrometry (MS/MS)

MS/MS adds another dimension to mass spectrometric measurement. Apart from improving the specificity of the technique, it can be used as a powerful structural elucidation tool. MS/MS instrumentation has been around for over 30 years and initially involved magnetic sector combinations and multiple quadrupole configurations, for example, the triple quadrupole was a popular choice until the mid 1990s. However, these have largely been superceded due to their relatively poor sensitivity and low mass resolution. Today there is a revolution unfolding regarding the combinations of mass analyzer technologies in MS/MS, in which the combined whole has an effect greater than the sum of its parts. These machines are sometimes referred to as hybrids reflecting some of the unusual combinations they involve (Cottingham, 2003). For proteomics studies the mass analyzer(s) must be capable of MS and MS/MS analysis; they must have the very best sensitivity and thus choices include TOF, FTICR, ion trap, quadrupole and linear quadrupole ion trap. All of these are true mass analyzers with varying properties and they have been combined in virtually all possible combinations (refer to Fig. 6). In addition there is a requirement in MS/MS to fragment the precursor species and for all mass ion analyzers, except FTICR, the choice is exclusively collision-induced dissociation (CID), also known as collisional activation (CA) dissociation. Depending on the type of mass analyzer CID can be conducted either under low collision energy (typically 2–100 eV or high collision energy (typically 500–10,000 eV). These two collision regimes give different types of fragment ions (Cooks, 1978) and the choice of analyzer is depend upon the specific application.

The revolution in high sensitivity MS/MS instrumentation started with the introduction of the quadrupole/TOF combination (Morris et al., 1996). Recent versions of this instrumentation have been described by other manufacturers (for example, see Baldwin et al., 2001). The quadrupole mass analyzer is usually restricted to a mass range of 2000–4000 u. Sobott et al. (2002) showed how this could be extended to 32,000 u by the development of a special quadrupole mass filter. This quadrupole-TOF MS/MS instrument is based on the orthogonal geometry (Guilhaus, 1995) which has proved to be a highly successful design. Sobott's TOF analyzer showed excellent sensitivity over the mass range. The quadrupole-TOF was one of the first MS/MS designs offering high sensitivity MS and MS/MS spectra with relatively straight-forward accurate mass in both MS and MS/MS modes; these features have led to its wide choice over the last decade.

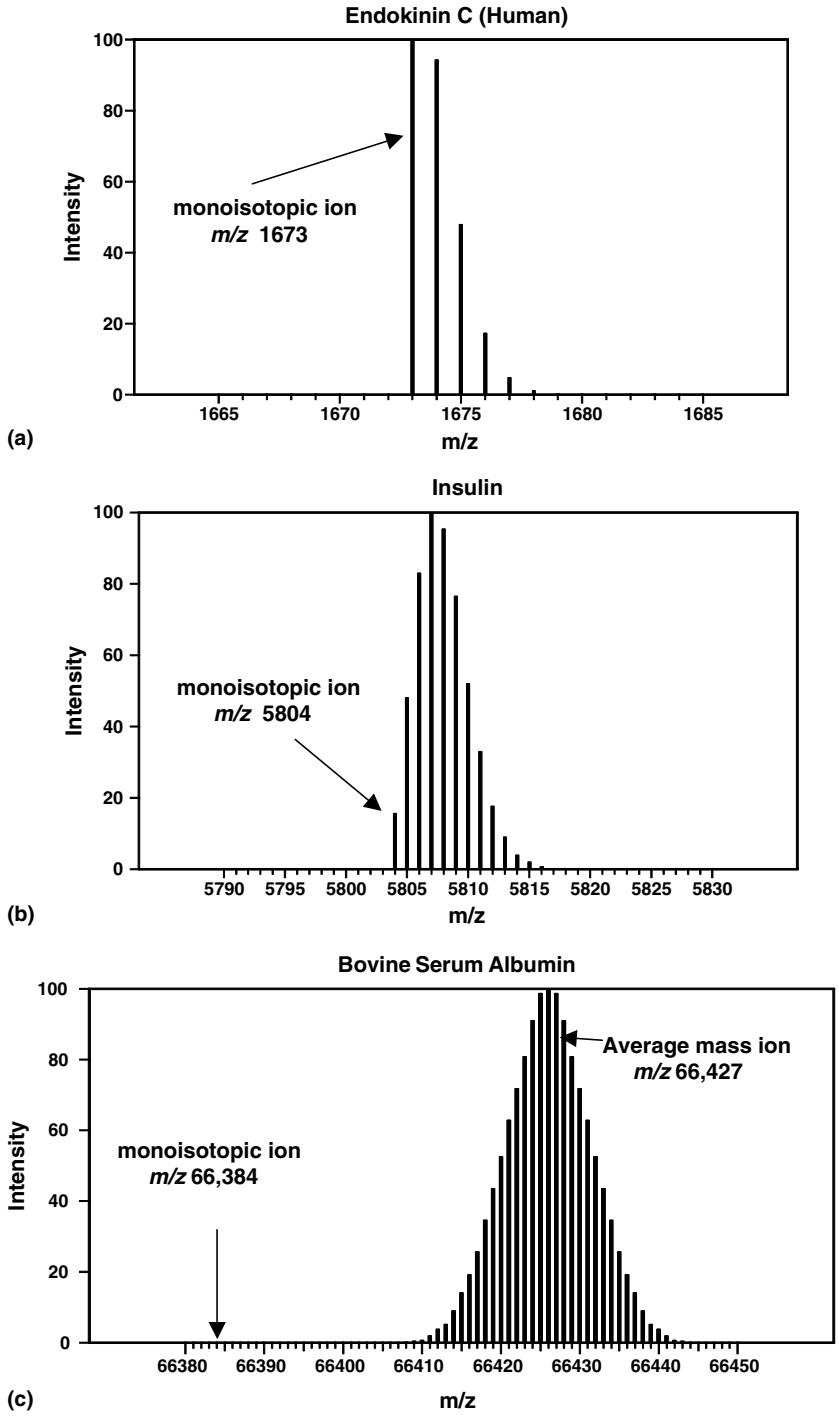


Fig. 4. Three theoretical isotopic patterns showing the relative intensities of the singly charged ion species. The height of the monoisotopic ion decreases as the size of the molecule increases: (a) endokinin C (human), m/z 1673; (b) insulin, m/z 5,804 and (c) bovine serum albumin, m/z 66,384.

5. Mass spectrometry in practice in proteomic studies

5.1. MALDI-TOF MS and peptide mass fingerprinting

For the MALDI analysis of a digested protein sample the mixture of peptides produced by trypsin or other protease action is mixed with a matrix solution. This matrix solution is a concentrated or saturated solution of a low molecular weight, UV absorbing, acid. In the case of peptides α-cyano-4-hydroxycinnamic acid (CHCA) is most commonly used as the matrix, both matrix and sample being dissolved in 50/50 0.1% trifluoroacetic acid/acetonitrile. However other matrices and solvents can be used for peptide ionisation; it has

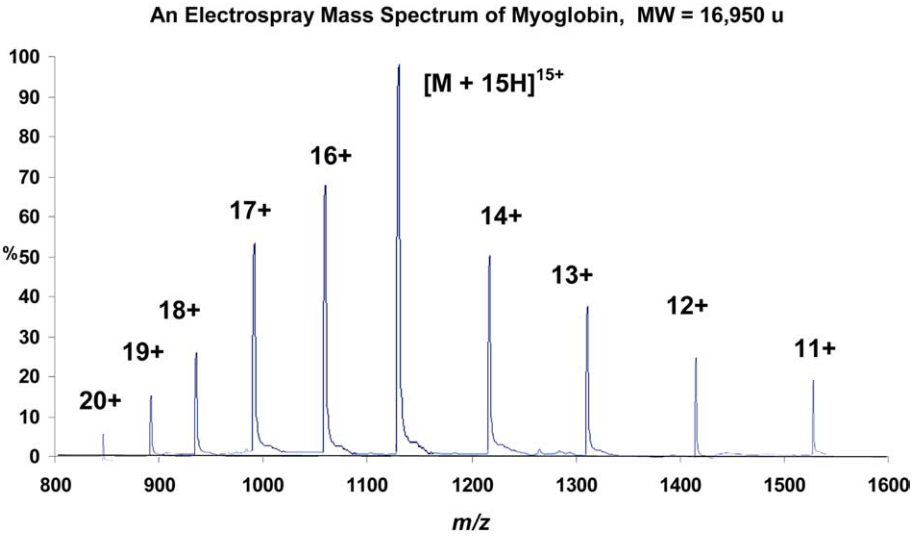


Fig. 5. An electrospray mass spectrum of Myoglobin (MW = 16,950) showing the multiply charging effect, the degree of charging depends also on the pH of the spray.

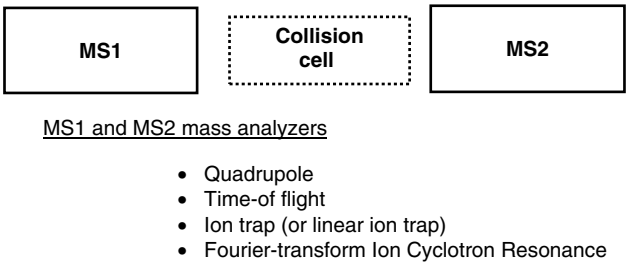


Fig. 6. Schematic diagram of an MS/MS spectrometer and the choices of analyzer that are available and suitable for high sensitivity proteomics studies. The analyzers for MS1 and MS2 can be of any type listed although certain combinations are preferred on the basis of performance and technical feasibility.

recently been suggested that a mixture of two matrices, CHCA and dihydroxybenzoic acid (DHB) gave improved results (Laugesen and Roepstorff, 2003).

Approximately 1 μl of the matrix/sample mixture is then spotted onto a MALDI plate and allowed to co-crystallise in a single spot. The matrix is in vast excess of the sample and absorbs UV light, thus when a UV laser is fired at the matrix/sample spot the UV energy is absorbed primarily by the matrix. This energy is passed from the matrix to the sample in a gentle manner in order that the ionisation process does not fragment the peptides; the effect of the laser energy is to produce a matrix plume which carries the analyte ions into a gas phase. While the detailed mechanism of MALDI is still under some discussion (Gluckmann et al., 2001), the overall effect of the process is to produce ions of the intact peptides without fragmentation. These ions are then separated by time-of-flight mass spectrometry in which the ions are accelerated from a fixed point and are allowed to drift through a TOF tube. Lighter ions (ions with a lower mass-to-charge ratio) travel faster down the TOF tube than heavier ions (with a higher mass-to-charge ratio) and so the time taken to travel down the TOF tube and reach the detector can be used to determine the mass-to-charge ratio of the ions. An advantage utilised in MALDI-TOF peptide mass fingerprinting is the high mass accuracy of the instrumentation.

For each firing of the laser therefore a mass spectrum is produced which contains significant ions relating to peptides from the digested protein together with ions from the matrix (usually at m/z below 500 Da) and usually also some from peptides arising from the autolysis of the protease used to digest the analyte protein. The first step of a peptide mass fingerprinting experiment is the deisotoping, by the spectrometer data processing software, of the mass spectrum produced. Isotope effects produce complex spectra because each peptide peak has a corresponding isotope peak 1 Da higher due to the naturally occurring C^{13} that is present in the peptide, and, dependent upon the mass of the peptide (i.e. the number of carbon atoms), successive isotope peaks of decreasing height will be observed at 1 Da difference further up the m/z range. Deisotoping has the effect of identifying these isotope peaks, which would otherwise cause false protein identifications, and removing them. Next a mass list is made from the mass spectrum so that the most intense peaks m/z values are recorded, this list is then usually compared to a filter which removes any recognised peaks that originate from the autolysis of the protease used for digestion. The peak list is then submitted to a database which compares the peptide masses from the spectrum with those of proteins theoretically digested and produces a list of possible matches and a scoring mechanism for their

likely correct identification. Before submission of the peak list the species from which the protein originated, the p*I* and approximate molecular weight of the protein, the protease used for digestion, any modifications made to the protein to prevent the reformation of the secondary structure, and the mass tolerance can be specified. Of these variables the protease used and the mass tolerance are the most critical. The protease used specifies how the theoretical proteins are digested in order to obtain the theoretical mass list for comparison with the spectra produced. The mass tolerance specifies how close in mass (measured in parts per million (ppm)) the submitted mass of a peptide and theoretical mass of a peptide must be in order to achieve a match. It has been reported in the past that a mass tolerance of 10 ppm is required to be confident of a peptide identification (Clauser et al., 1999). As noted before MALDI-TOF MS allows high mass accuracy with specifications of 80 ppm with external calibration (when the calibrant is on a separate part of the MALDI plate) and 10 ppm with internal calibration (when the spectrum is calibrated using peaks within the spectrum itself).

A number of methods of obtaining sufficient mass accuracy for accurate peptide and protein identification are possible. The first is the addition of a number of internal standard peaks of known mass which can then be used to calibrate the spectra. The second is the use of the autolysis peaks that originate from trypsin acting upon itself to produce a number of peptides of known mass. These trypsin autolysis peaks have been well characterised (Harris et al., 2002) and therefore can be used to calibrate the mass spectra produced. The third option is known as close external calibration. This involves the calibration of the mass spectrometer by a calibration sample that is positioned as close to the unknown sample as possible, the created calibration then being used to run the uncalibrated sample. The mass spectrum produced has slightly less mass accuracy (in our facility specifications from a Voyager DE-STR are 80 ppm for close external calibration compared to 10 ppm for internal calibration) and this is submitted to the database in a slightly different manner. Firstly, the data are submitted and compared with theoretical proteolytic peptides with a high mass tolerance (e.g. 250 ppm). This returns a large number of possible protein identifications and from these the best fit is chosen and the theoretical masses of its peptides are used to calibrate the spectrum suspected to be of the peptides from this protein. This recalibrated spectrum is then resubmitted to the database with a mass tolerance of 10 ppm. If the ''best fit'' protein was correct then all the peptide peaks would have been calibrated accurately and give the 10 ppm mass tolerance required. If however the ''best fit'' protein was incorrect then the calibration of the spectrum would not have improved the mass accuracy of these peaks and therefore the 10 ppm search would fail. Of these methods the use of trypsin autolysis peaks as an internal standard is the most common when applicable as the autolysis peaks should always be present in the spectra produced and have been detected even with trypsin modified to reduce autolysis.

An example of a peptide mass fingerprint experiment is shown in Fig. 7. Fig. 7 shows the mass spectrum produced by the analysis of a spot picked from a 2D SDS–PAGE gel following digestion with trypsin the peaks marked with • represent trypsin autolysis peaks that were used to internally calibrate the mass spectrum. The peaks marked with ■ represent peaks identified as peptides produced by the digestion of phosphoglycerate kinase, these peptides are listed with their actual masses in Table 6. It should be noted that the peaks in Table 6 (m/z submitted column) do not match those seen in Fig. 7. This difference arises as Fig. 7 shows the mass spectrum before deisotoping has been performed and so the importance of this step is illustrated. In assessing mass accuracy it can be seen that the experimental data and the theoretical masses of the peptides differ by less than 10 ppm. Eleven peaks were matched to this protein representing 27% of the protein sequence. It can be seen in Fig. 7 that there are peaks in the spectrum that are not identified as originating from this protein; these may arise from two proteins being picked in a single spot from the gel, unknown trypsin peaks (such as that at m/z 1060 which occurs in a standard trypsin sample but which has yet to be identified), matrix adducts, missed cleavages during digestion, or contamination of the sample or matrix solution.

After the calibration of spectra the most important factor in determining correct protein identification is the quantity of the protein applied to, and the efficient removal of the protein from, the gel before analysis by MALDI. With the more sensitive silver staining technique (and therefore the least amount of sample), and assuming a loss of 90% of the sample upon extraction before MALDI analysis, the sensitivity of the MALDI-TOF (typically ca. 5 fmol) should be sufficient to identify the protein (Shevchenko et al., 1996). Experimentally however problems have occurred in identifying proteins which have been identified by the less sensitive coomassie blue staining procedure. It can be concluded that either the sample is being lost upon extraction or factors such as SDS are interfering with the MALDI ionisation of the sample. A possible solution to this problem is the clean up and concentration of the samples prior to MALDI and after picking and digestion. This can be achieved by microscale solid phase extraction procedures using commercially available pipette tips packed with stationary phase packing such as ZipTips® from Millipore. These tips allow the peptides to be retained upon the stationary phase of the tip whilst salts are washed off, the peptides then being eluted from the tip directly onto the MALDI plate for analysis.

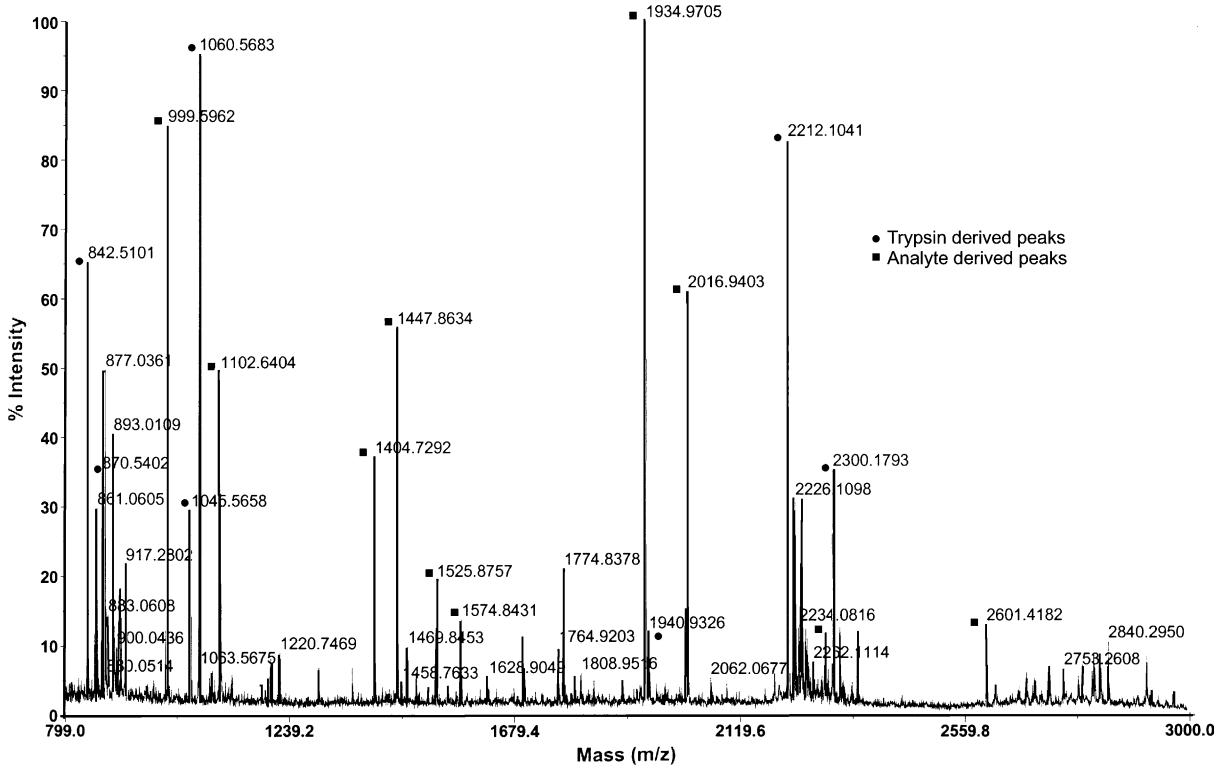


Fig. 7. A peptide mass fingerprint experiment: mass spectrum produced by the analysis of a spot picked from a 2D SDS–PAGE gel following digestion with trypsin. The peaks marked with • indicate trypsin autolysis peaks that were used to internally calibrate the mass spectrum. The peaks marked with ■ indicate peaks identified as peptides produced by the digestion of phosphoglycerate kinase.

Table 6
Identification of phosphoglycerate kinase

m/z submitted	*m/z* matched	Delta (ppm)	Sequence of peptide
999.5959	999.5991	3.1855	(K) FSLAPLVPR (L)
1102.6402	1102.6373	2.6429	(K) RPFAAIVGGSK (V)
1220.7472	1220.7479	0.6248	(K) VILSTHLGRPK (G)
1404.7299	1404.7375	5.3725	(K) ELDYLVGAVSNPK (R)
1447.8636	1447.8677	2.8167	(K) FLKPSVAGFLLQK (E)
1525.8751	1525.8841	5.9406	(K) GVSLLLPTDVVVADK (F)
1573.8428	1573.8437	0.6144	(K) GVTTIIGGGDSVAAV- EK (V)
1933.9712	1933.9772	3.0853	(K) LASLADLYVNDAFGTAHR (A)
2015.9488	2015.9522	1.6474	(R) ADLNVPLDDNQTITDDTR (I)
2285.1848	2285.1811	1.5930	(K)VGVAGVMSHISTGGG-ASLELLEGK (V)
2600.4121	2600.4035	3.3117	(K) AQGLSVGSSLVEEDKLELATELLAK (A)

5.2. Electrospray tandem mass spectrometry (ESI-MS/MS) sequencing of peptides for protein identification

The second method of mass spectrometry used for protein identification is electrospray tandem mass spectrometry. This process again analyses the peptides produced by the trypsin hydrolysis of the proteins in order to identify the protein from which they originated. The mass spectrometer is usually coupled to a HPLC system, although good quality data have also been achieved from the direct infusion of the entire sample. Electrospray ionisation allows the analysis of ionised molecules in solution and is therefore ideal for combination with HPLC separations. During electrospray ionisation a sample solution or HPLC eluent enters the ion source as a fine mist of droplets via a needle which has an accompanying flow of nitrogen gas surrounding it. A high voltage is applied to the needle through which the solution arrives in the source causing the droplets produced to be charged on the surface. This droplet is then desolvated, mainly due to the presence of the nitrogen gas, and its diameter therefore becomes reduced. This process continues until a given point, referred to as the Rayleigh stability limit, at which the electrostatic

repulsion of the ions present becomes greater than the surface tension holding the droplet together. This whole process results in the ions being released from the liquid droplet to produce gas phase ions that are drawn into the mass spectrometer. Again the mechanism of electrospray is under some debate as seen by past reviews into the understanding of the process (Bruins, 1998; Gaskell, 1997).

During LC-ESIMS/MS the HPLC is used to separate the peptides produced by digestion prior to their entering the mass spectrometer. Commonly miniaturised HPLC systems are now used for peptide analysis which use capillary columns (300 μm internal diameter) or nanocolumns (75 μm internal diameter). These are used at low flow rates, for example nanocolumns are used at approximately 200 nl per minute, but still allow large injection volumes through multidimensional HPLC (see later in this review). As the electrospray mass spectrometer is a concentration-dependent detector this reduction of the flow rate involved in the separation allows great increases in sensitivity with limits of detection of around 5–10 fmol being reported (Hoyes et al., 2002). Usually C18 columns are used for peptide separations and HPLC runtimes are approximately 45 min to an hour. New columns have recently been released that use monolith material as their stationary phase and allow faster separations of approximately 15 min to be used (Premstaller et al., 2001; Walcher et al., 2002). Once the peptides enter the mass spectrometer they are subjected to tandem mass spectrometry. In this process the molecular ion of the peptide is chosen as the ion of interest this ion is then fragmented by collision with a collision gas to produce a number of fragment ions. The m/z ratios of these fragment ions are then recorded to give a fragmentation spectrum of the peptide. The peptide can be viewed as an amino acid chain that be fragmented at any point, however fragmentation occurs preferentially across the nitrogen and carbon bond between the two amino acids. This fragmentation can produce two fragment ions, a y-ion is formed when the charge is retain on the C-terminus of the peptide and a b-ion is formed when the charge is retained on the N-terminus of the peptide (see Fig. 8). Alternative ions

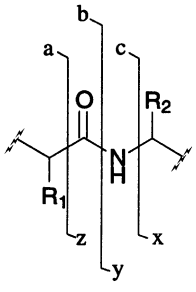


Fig. 8. Preferential fragmentation between the nitrogen and carbon bond of two amino acids in a peptide or protein chain.

such as a, c, x and z ions can also be formed from fragmentations as indicated, but these are less prevalent than the b and y ions. If the peptide fragmented is singly charged the charge is retained upon either the C or N terminus and therefore either the y or b ions are seen, respectively. If however the peptide is doubly charged then it is possible that both termini of the peptide remain charged and both the b and y ions are seen in the fragmentation spectrum. This spectrum can be further complicated if the peptides are triply charged, as the fragment ions produced can be doubly or singly charged as well. Rather complicated tandem mass spectra (MS/MS spectra) can therefore be produced by the fragmentation of a relatively simple peptide. The interpretation of the MS/MS spectra in order to obtain the peptide sequence is therefore a long laborious and difficult process even with modern software. It should be noted however that if a suspected protein of known sequence is being analysed the application of manual or software assisted sequencing of the peptide may be appropriate.

The next consideration is how are the peptides to be fragmented in the mass spectrometer selected, and how the MS/MS spectra used to identify the protein picked from the gel are generated? Acquisition of data during HPLC MS runs of peptides is usually performed using data-dependent acquisition (DDA). A threshold of the background noise detected from the HPLC solvents is set and if necessary an exclusion list of known background peaks can be made so that these are not analysed during the DDA run. The mass spectrometer is then programmed to perform full scan mass spectrometry and then tandem mass spectrometry on any ion that exceeds that threshold and repeat the fragmentation until the ions abundance is below the threshold. Before fragmentation it is possible for the charge of the peptide to be determined by measuring accurately the difference in mass between the peptide ion and its C^{13} isotope ion. This naturally occurring ion appears one mass unit higher than the molecular ion. If the peptide is singly charged (and therefore z is one) the ion will be one mass/charge unit higher, if it is doubly charged (z is two) then the ion will be half a mass/charge unit higher than the molecular ion. This is useful information as a doubly charged ion when fragmented will produce b and y ions with one charge each and some of these will occur higher up the m/z scale than the molecular ion. The m/z range to be scanned is therefore automatically decided dependent upon this information. Once complete the resulting file contains a series of full scan mass spectra and MS/MS spectra of any selected ions. Fig. 9 shows the total ion current of a DDA run of a cytochrome c digest and an ion at m/z 729. From a run of this sort the molecular ions m/z value, the charge on the peptide and the MS/MS spectra are used for identifying the protein by a database. Firstly, the peptide mass is used to select

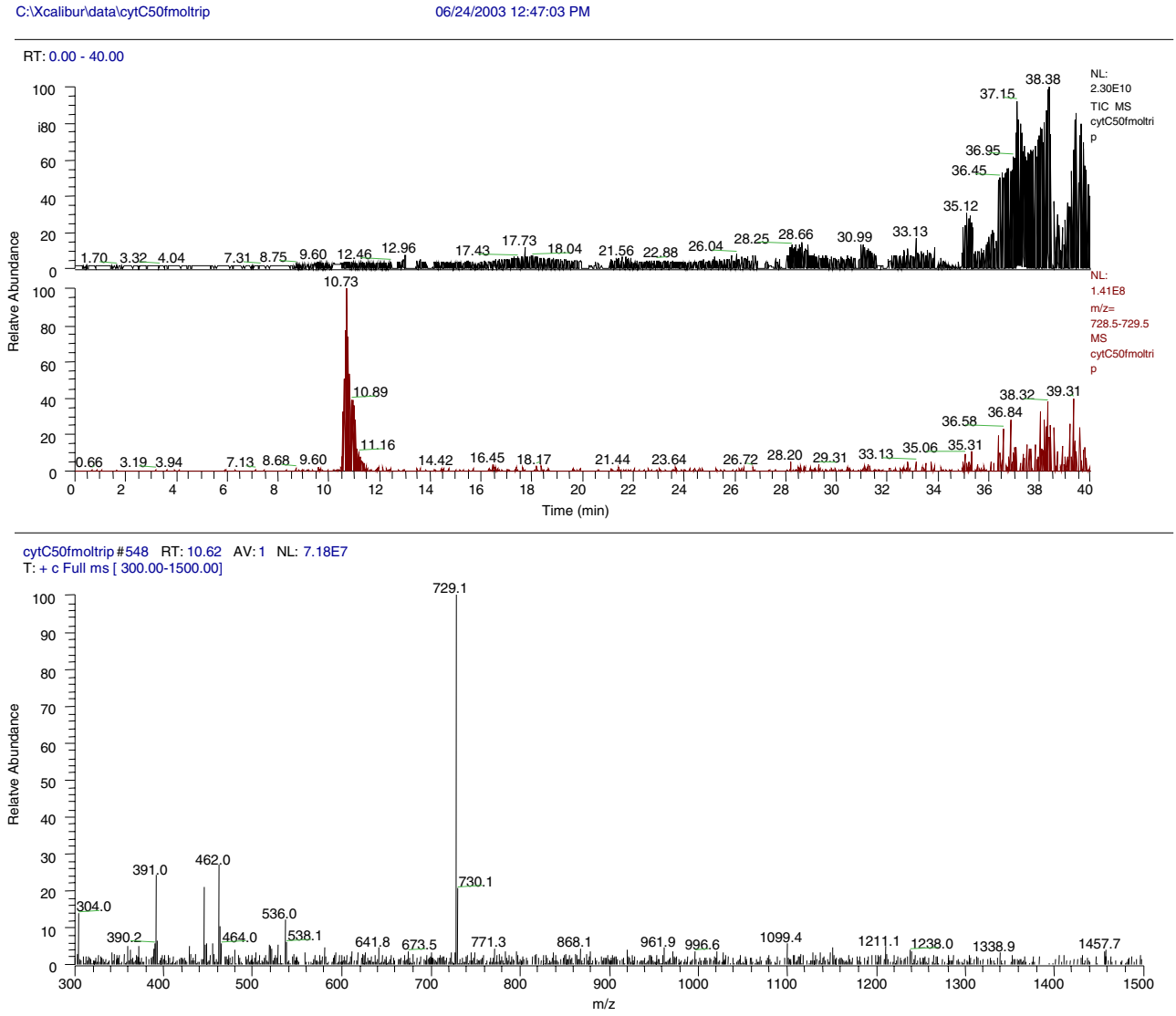


Fig. 9. Total ion current, extracted ion chromatogram and full scan mass spectrum of m/z 729 from a DDA run of a cytochrome *c* digest.

all peptides of the same mass as that detected in the run (using both the observed molecular ion and its charge to determine the actual mass of the peptide). This provides a list of candidate peptides one of which should be the peptide analysed. These peptides are then theoretically fragmented to generate the m/z ions expected if the candidate peptides were fragmented. These fragment ions are compared with those actually observed in the DDA run and the peptides scored as to how well their theoretical data match the observed spectrum. An example of the correct identification of the sequence of the cytochrome *c* digest peptide at mass 729 is shown in Table 7/Fig. 10 which shows the observed fragmentation spectrum, the theoretical fragments of a cytochrome *c* peptide with the matching ions being highlighted. This process is performed for every MS/MS data file in the DDA run and as only peptides arising from one protein should be in the sample then the protein should be identified by a number of its peptides fragmentation spectra being matched to the observed data.

An alternative fragmentation experiment, known as an Ion Tree experiment, is available when using ion trap mass spectrometers. Ion traps "trap" ions and therefore more complex experiments can be performed, enabling their use in sequencing a peptide. The most important aspect of the ion trap in relation to the Ion Tree experiment is the ability to fragment the peptide ion to form fragment ions that can themselves be isolated individually, then fragmented further. This allows the analysis of fragment ions from the MS/MS spectra fragments thereby creating an MS^3 experiment. In the Ion Tree experiment the peptide is fragmented, the fragment ions of greatest intensity are then subsequently fragmented, and the MS^3 spectra of these ions recorded. The latter data are then used to produce clearer mass spectrometric data and consequently better sequence

Table 7
b and y Ions for cytochrome *c* peptide TGQAPGFSYTDANK

Sequence	#	b	y	(+1)
T	1	102.1	1456.7	14
G	2	**159.1**	1355.6	13
Q	3	**287.1**	1298.6	12
A	4	**358.2**	**1170.5**	11
P	5	**455.2**	**1099.5**	10
G	6	**512.2**	**1002.5**	9
F	7	**659.3**	**945.4**	8
S	8	**746.3**	**798.4**	7
Y	9	**909.4**	**711.3**	6
T	10	**1010.5**	**548.3**	5
D	11	**1125.5**	**447.2**	4
A	12	**1196.5**	**332.2**	3
N	13	**1310.6**	**261.2**	2
K	14	1438.7	**147.1**	1

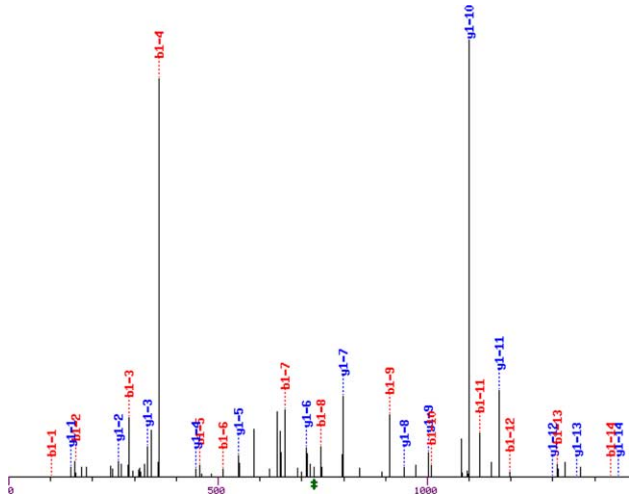


Fig. 10. Identification of fragment ions from the DDA tandem mass spectrometry fragmentation of the ion at m/z 729 that match those of the cytochrome *c* peptide TGQAPGFSYTDANK.

determination: while the MS/MS fragmentation spectra can contain both b and y ions making interpretation of the data sometimes difficult, in the ion tree experiment the MS^3 spectra of a prominent b ion will contain fragments from the b series of ions alone and similarly for the MS^3 of the y ion. Therefore, the MS/MS and MS^3 spectra are recorded and compared to determine which MS/MS ions belong to which series (b or y) thus making interpretation of the data easier (Zhang and McElvain, 2000).

Alternatives to performing a HPLC separation prior to the MS/MS fragmentation of the peptides have been proposed in order to increase the throughput of this identification procedure. The first is the application of the sample into a static nanospray needle from which the sample is electrosprayed directly as a mixture of all the peptides without separation (Sheil et al., 2001). The second is the trapping of the peptides at a high flow rate on a "trap" column which is then eluted at a low flow rate, in electrospray friendly conditions with no separation of the peptides (Witters et al., 2003). Of these the first is effective for very low concentrations of sample but requires a lot of manual labour and a relatively long runtime, the second allows for higher throughput of samples, with 300 samples a day being analysed. Both methods rely upon the DDA run being able to pick more than one ion to fragment, this being referred to as dynamic exclusion. Under dynamic exclusion conditions again the ions above a threshold are chosen, usually in a DDA run only the most intense ion is fragmented. During a dynamic exclusion DDA run firstly the most intense ion is fragmented, this ion is then ignored when selecting an ion to fragment and so the second most intense ion is selected for fragmentation and this process is then repeated to allow the fragmentation of as many of the peptide ions as possible. An alternative approach to the comparison of the MS/MS data to that of theoretical peptides is a search based on a "sequence tag". The MS/MS data are studied (usually with the aid of sequencing software) in order to interpret at least part of the data and produce part of the sequence from that peptides MS/MS data. This sequence is known as a sequence tag and provides a part of the amino acid sequence of the analysed peptide. This sequence tag is then entered into a database and any peptide containing the same order of amino acids within its peptide sequence is determined. This process differs from the usual searching of MS/MS data in the fact that one set of data is interpreted alone (and usually more experimenter influence can be exerted) than with DDA runs. The advantage and disadvantage with this process is that the experimenter has more influence over how the data are interpreted. If an ion has been mistakenly chosen as important then it can be ignored, however error of judgement is also more likely.

5.3. MALDI versus electrospray

The two methods of protein identification by mass spectrometry outlined above offer different advantages and disadvantages. MALDI-TOF analysis is easy to automate and allows the analysis of large numbers of samples in a short period of time (for example a plate of 96 samples can be run in less than an hour). The protein identification relies purely upon the matching of the peptide masses accurately. In samples of low concentrations or high levels of contaminant from workers or the gels themselves many possible proteins can be equally well identified as possible correct hits. Therefore, although the process is high throughput, care needs to be taken to desalt the samples and accurately calibrate the spectra produced. ESI-MS/MS offers more certainty as to the protein identification as the peptide sequence as well as its mass is used to identify the peptides in the

sample. The runtime of the HPLC run however limits the throughput achievable. Even when a trap-elute system is used as previously described above a runtime of 5 min per sample is required and fewer peptides are matched compared to the full HPLC run and MALDI analysis. Little difference between the two methods has been reported, however LC-ESI-MS/MS allows automated desalting of samples which is only just being made easier for MALDI with the release of ZipTips in 96 well format for direct cleanup and elution onto MALDI plates. When comparisons have been made in the past between the two mass spectrometric methods for protein identification it has been concluded that the two methods complement each other (Lim et al., 2003; Bodnar et al., 2003) and that "the preference for one or the other in a protein chemistry lab is not an option" (Marko-Varga et al., 2003). The advent of MALDI TOF/TOF instruments, with two time of flight tubes allowing collisional dissociation experiments analogous to ESI-MS/MS fragmentation of selected ions, and of post-source decay/delayed extraction experiments, allow additional fragmentation data from MALDI systems.

A large number of Internet sites have been established which can be used to identify proteins using both MALDI peptide mass fingerprinting (PMF) data and LC-ESI-tandem mass spectrometric data (MS/MS). Some sites also provide useful tools such as theoretical digests of known proteins or amino acid sequences and example protocols for use in proteomic research. Table 8 lists some examples of useful sites currently available.

5.4. Alternatives to SDS–PAGE for protein analysis

Both of the described methods require that the proteins are identified by the staining procedure in the SDS–PAGE gel, efficiently extracted from the whole gel, digested and the peptides removed from the excised gel piece. This process has a number of disadvantages. In order for the protein to be chosen it must be visible by the staining process used and so proteins in concentrations below the limit of detection of the staining procedure are not selected. The efficiency of the extraction of the protein from the gel can vary and may result in the loss of sample, also each gel may take a number of days to set-up and run followed by spot picking and digestion leading to a lengthy preparation time prior to mass spectrometric analysis. Due to these factors a number of researchers have considered the replacement of the SDS–PAGE step with HPLC separations. Separation of proteins from whole cell lysates has been performed by reverse phase (RP) chromatography for studying protein changes during cancer progression (Chong et al., 1999): however most protein separations utilise more than one type of column. The separation of intact proteins has been studied using multidimensional HPLC in which the sample is applied to a first column and eluted onto a second column allowing different stationary phases and therefore different separation mechanisms to be combined and facilitate greater resolution than a single HPLC column alone.

The separation of intact proteins has been studied by a number of researchers. Generally the second dimension (i.e. the second column in line) is a reverse phase column with the difference being in the column used for the first stage of the separation. First dimension columns used have included size exclusion columns (Opiteck et al., 1998), affinity columns (Coutre et al., 2000) and isoelectric focussing columns (Wall et al., 2000; Lubman et al., 2002). In most of these methods the fractions require collection and digestion before the MS analysis to determine protein identity could be performed. One group however used the direct coupling of a multidimensional separation to an ESI mass spectrometer in order to study the molecular weights of the proteins separated (Liu et al., 2002). Alternatively columns packed with immobilised trypsin are now available and have been applied for the on-line digestion of proteins (Blackburn and Anderegg, 1997; Wang and Regnier, 2001). Most work utilising multidimensional HPLC however has involved the digestion of the proteins in the whole sample. Mul-

Table 8
Internet sites used for proteomic analysis

Internet site	Useful utilities	Site address
Protein prospector	Search databases with PMF and MS/MSdata. Provide theoretical data for a known peptide sequence	http://prospector.ucsf.edu
PROWL	Search databases with PMF and MS/MS data. Provide theoretical data and useful protocols	http://prowl.rockefeller.edu/PROWL/prowl.html
ExPASy	PMF database searches, theoretical tools and links to many other sites	http://us.expasy.org
PeptideSearch	Search PMF data or a sequence tag	http://www.mann.embl-heidelberg.de/GroupPages
Mascot	Search databases with PMF and MS/MS data and sequence tags	http://www.matrixscience.com
PepMAPPER	Search databases with PMF data	http://wolf.bms.umist.ac.uk/mapper/
MOWSE	Search databases with PMF data	http://srs.hgmp.mrc.ac.uk/cgi-bin/mowse
PredictProtein server	Gives theoretical fragments and data on a protein amino acid sequence	http://www.ebi.ac.uk/~rost/predictprotein/
BLAST	Search sequence tags	http://www.ncbi.nlm.nih.gov/BLAST/

tidimensional separation of the peptides is then performed prior to ESI-MS/MS analysis or MALDI analysis rather than the separation of the individual proteins. This chromatography is given the acronym MudPIT for multidimensional protein identification technology (Wolters et al., 2001). ESI-MS/MS analysis is more commonly used as MALDI analysis requires the collection of fractions prior to analysis. However, with the low flow rates now available with miniaturised HPLC this fraction collection can be done directly onto the MALDI plate, making the process of HPLC-MALDI analysis more user friendly. Due to the increased complexity of the sample after digestion, with each individual protein being digested to produce a number of peptides, even more resolution is required prior to analysis.

Miniaturised multidimensional HPLC has been used to separate complex mixtures of peptides from a number of sources ranging from *Arabidopsis* chloroplasts to bovine microtubule preparations and ribosomal complexes (Schwartz et al., 2002; Wagner et al., 2003). These separations usually involve either material packed into a picofrit column (McDonald et al., 2002) or a nano-reverse phase analytical column with one or more trap columns before this (Schwartz et al., 2002). These trap columns are very short (5–15 mm) columns of usually 300 μm internal diameter that have sample loaded onto them at relatively high flow rates (30 μl per minute). Such traps are also used to load large volumes onto nanocolumns by firstly injecting the sample onto the trap column and then washing off the trap column at a much reduced flow rate onto the nanocolumn. When performing multidimensional HPLC with these trap columns a number of valves are used to switch columns in and out of line. The example in Fig. 11 has been used to study the *Arabidopsis* chloroplast proteome in which 436 proteins were identified from the entire multidimensional HPLC-MS/MS run (Schwartz et al., 2002). The sample is injected into a loading pump flow and passes through a strong cation exchange (SCX) trap column (on valve A) followed by a C18 (RP) trap column (on valve B) and then to waste. Meanwhile the gradient pump flows into valve B and directly over the analytical nanocolumn. Some peptides are retained on the SCX trap due to their being positively charged whilst those that do not are retained on the RP trap (this trapping also serves to de-salt the sample). The peptides positive charge can arise from the protonation of the nitrogen atoms within the sidechains of the arginine or lysine amino acid residues at which the trypsin breaks the peptide. Valve B (see Fig. 11) is then switched so that flow from the gradient pump flows over the RP trap and then onto the nanocolumn. A gradient of increasing acetonitrile is then pumped from the gradient pump and this elutes the RP trapped peptides onto the nanocolumn where they are separated and elute into the mass spectrometer ESI source. After the nanocolumn separation has been performed and the nanocolumn gradient returned to its starting conditions valve B is switched again to take the nanocolumn out of line from the two trap columns. Next a low molarity salt "plug" (e.g. 10 mM NaCl) is injected onto firstly the SCX followed by the RP trap. This plug elutes the least well bound peptides from the SCX trap and these are then retained on the RP trap. Valve B is then switched again as before and the RP retained peptides separated and analysed. This procedure is then repeated with increasing molarity plugs which elute progressively more tightly bound peptides. This process fractionates the SCX retained peptides and these are then separated on the analytical column before analysis by the mass spectrometer. Such multidimensional HPLC allows the analysis of very complex mixtures of peptides. A study compared

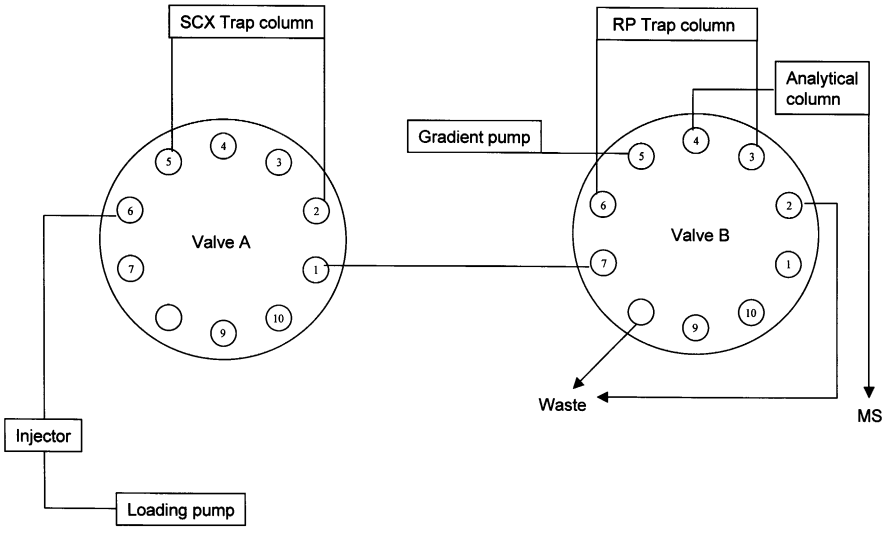


Fig. 11. Multidimensional HPLC: use of a number of valves to switch trap columns in and out of line.

multidimensional separations with varying numbers of columns in line and found that the combination of two traps and one analytical column resulted in 66 proteins being identified from an extraction of bovine microtubule associated proteins compared to 26 with the use of one analytical column (McDonald et al., 2002). Multidimensional HPLC has also been reported to have the further advantage that it exhibits near quantitative recovery of the peptides separated (Davis et al., 2001). Although multidimensional HPLC allows the separation of whole proteomes before on-line analysis it does require very long run times, with the average time required for the whole experiment being 20–25 h. However a much larger proportion of the total proteome is studied rather than just the proteins visible on the SDS–PAGE gel, and the run time of the experiment is less than that required for the preparation and running of a SDS–PAGE gel.

5.5. The study of differential protein expression

A common argument in favour of SDS–PAGE use is that comparisons can readily be made between two gels and thus proteome differences detected. However multidimensional HPLC has successfully been applied to the analysis of differences in protein expression (Opiteck et al., 1998) with the omission of electrophoresis. Proteins from *Escherichia coli* were separated by size exclusion chromatography followed by reverse phase chromatography, with online monitoring of their UV absorbance, and fraction collection. The same sample (after inoculation with a plasmid containing the gene for the SH2 domain of $pp60^{c\text{-}src}$) was then run and the two UV traces compared and the difference seen determined to be due to SH2 by mass spectrometric analysis of the collected fraction.

An alternative is to determine the relative amounts of proteins in samples being compared by using isotope-coded affinity tag (ICAT) experiments (see Fig. 12). The two protein samples are firstly labelled with one of two tags. To one sample a tag with a thiol reactive group which selectively targets cysteines within a proteins amino acid sequence is added which contains eight hydrogen atoms. This tag also contains a biotin moiety for selective retrieval of the labelled peptides. To the second sample a similar tag which contains eight deuterium instead of the hydrogen atoms is added. The two samples can then be combined in equal amounts and digested to produce the peptides from the proteins from both samples. Identical peptides are thus produced from the two samples, both of which are labelled, one with the hydrogen tag and the other with the deuterium tag. These identical peptides in the two samples now have an eight mass difference due to the one mass unit difference in mass between hydrogen and deuterium atoms. This peptide mix is then passed through an affinity cartridge

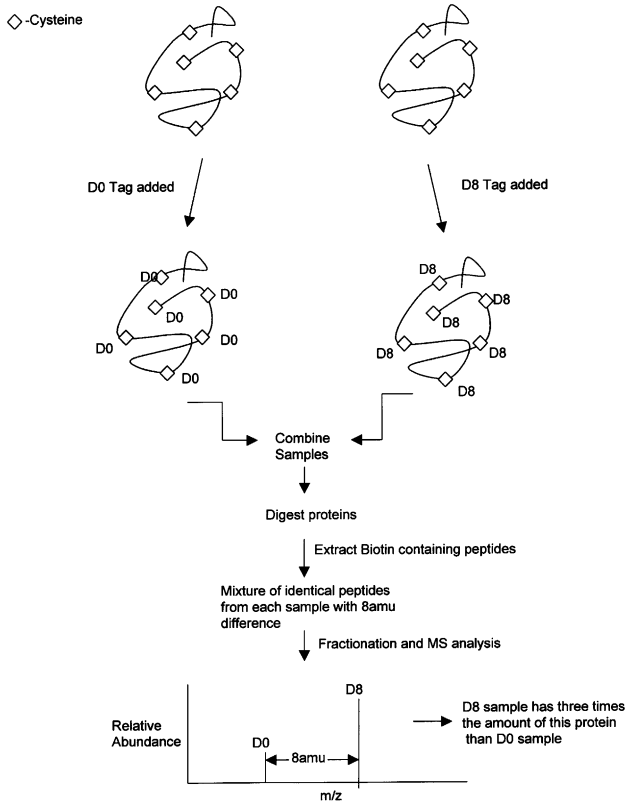


Fig. 12. Quantitation: determination of the relative amounts of proteins in samples by use of isotope-coded affinity tag (ICAT) experiments.

which contains an avidin stationary phase that will bind the tags biotin moiety and therefore retain all the labelled peptides. Next these are eluted from the cartridge and the sample is analysed by mass spectrometry and may be fractionated to reduce sample complexity if required. The spectra produced have ions representing a mixture of these coupled peptides from the two samples that differ by 8 amu. The relative height of these peaks to each other represents their comparative ratio in the two samples. The peptides that differ significantly can then be sequenced by LC-ESI-MS/MS or post-source decay analysis within the MALDI TOF and be used to identify the protein from which they originated (Griffin et al., 2001). The HPLC separation required in order to resolve the pairs of peptides for quantitation of complex mixtures has been developed (Lee et al., 2003) and advances have been made in order to make automated MALDI interpretation of such experiments easier (Applied Biosystems, 2002). The expense involved in performing ICAT experiments however is still a limiting factor and it has been claimed that neither ICAT or 2D SDS–PAGE provide comprehensive coverage on a proteome wide scale (Patton et al., 2002).

A major factor that limits the proteome coverage yielded by ICAT analysis is the requirement of a cysteine amino acid in the proteins amino acid sequence for

the addition of the ICAT labels. It has previously been stated that one protein in seven does not contain cysteine (Rabilloud, 2002) and therefore other methods have been developed to overcome this difficulty. One such method is termed mass-coded abundance tagging (MCAT), which on a similar principle to ICAT, however, in the MCAT process the terminal lysine residues (derived from trypsin digestion) are either left unmodified or converted to homoarginine through guanidination (Hale et al., 2000). The peptides when mixed now have a mass difference of 42 amu, which can then be compared by LC/MS or MALDI. A drawback of the MCAT process is the difference in ionisation efficiency caused by the guanidination, which may affect the ratio of the observed peaks (Hamdam and Righetti, 2002). In order to further increase the number of peptides labelled a number of other tagging experiments have been devised and termed global internal standard strategies (GIST). The first of these involves the labelling of the primary amino group of tryptic peptides with either *N*-acetoxysuccinimide or *N*-acetoxy-[$^2H_3$]succinimide (Chakraborty and Regnier, 2002): however it was noted that, depending on the amino acid sequence, the peptide could be labelled once or twice and so different mass shifts were seen. In order to overcome this problem the experiment was repeated using a light and heavy (deuterated) succinic anhydride to determine the number of labels; this reduced the problem but could not remove it completely (Wang et al., 2002). A second approach involved the use of stable isotope labels that non-specifically label amino acids (Yao et al., 2001). In one experiment yeast was grown on natural nitrogen and medium enriched with ^{15}N and the differential analysis of 42 high abundance proteins was examined (Oda et al., 1999). In a second experiment peptides from two samples were methylated with d_0-methanol and d_3-methanol, respectively, prior to mixing and analysis conducted in a similar manner to that used in ICAT (Goodlett et al., 2001).

5.6. The study of protein modifications

Mass spectrometric analysis of the proteome can also be used to analyse protein modifications. Phosphorylation is the most commonly observed and has been most extensively studied (Kalume et al., 2003). Again SDS–PAGE was, until recently, the method of choice for studying phosphorylation. In these experiments radiolabelled phosphate would be added to a sample and the radiolabel incorporated into any phosphorylation sites of the proteome. The sample would then be run on a SDS–PAGE gel and the radiolabelled, phosphorylated proteins identified by their detection using phosphorimaging (Bykova et al., 2003). Alternatives now exist for studying the phosphorylation of proteins. The first option is to selectively extract phosphorylated proteins or peptides. This can be performed using immobilised metal-ion affinity chromatography (IMAC) in which a supporting stationary phase is "loaded" with a metal ion such as iron. This metal loaded IMAC column then has an affinity for the phosphate groups of the phosphoproteins or phosphopeptides and therefore can be used to purify these, followed by their identification as described earlier. A number of reviews covering the development and utilisation of IMAC separations have been produced (Chaga, 2001; Gaberc-Porekar and Menart, 2001) and the development of on-line systems for the purification of phosphopeptides and direct analysis have been described (Cao and Stults, 1999). An alternative method is again to digest the entire sample and use the mass spectrometer to selectively fragment or highlight phosphorylated peptides. This process relies on the fact that under fragmentation conditions the peptide ion will lose its phosphate group to produce a specific fragment. This fragment arises due to the loss of H_3PO_4 from the phosphorylated peptide and produces a fragment ion 98 amu less in singly charged peptides, 49 amu in doubly charged peptides and so on. Analyses have been performed in which the peptide mixture is separated by nano-HPLC and eluted into the mass spectrometer. For each peptide the molecular ion is fragmented with low collision energy. This fragmentation causes the loss of the phosphate group if the peptide is phosphorylated and a loss as described is detected, if however the peptide is non-phosphorylated then no such loss is seen. Upon the observation of these losses with low collision energy the mass spectrometer switches to perform a full tandem mass spectrometry experiment and so only phosphorylated peptides under go tandem mass spectrometry. This therefore allows the specific identification of proteins have that been phosphorylated in any given sample (Bateman et al., 2002). Alternatively the tandem mass spectrometric spectra can be searched for these losses from the molecular ion after all the fragmentation spectra have been recorded from within the fragmentation spectra of both non-phosphorylated and phosphorylated peptides (Scigelova et al., 2002). The phosphorylated peptides MS/MS spectra are highlighted and can be selected from within the entire experiment and searched against databases on their own.

5.7. Investigation of protein–protein interactions

An important aspect of proteomics is how proteins interact with each other, forming complexes of proteins in doing so. There are a number of ways in which this has been studied in the past and these are described first of all. The first is that the protein, and its complexing proteins, are immunoprecipitated from solution. These proteins are then separated by 1D SDS–PAGE, the identity of these proteins is then established by immunoprecipitation with a precipitant specific for the

suspected protein that is undergoing interaction. In the second method, again the suspected interacting protein identifications are known. The genes encoding these proteins have part of two genes added to them so that in the result of them "interacting" the two gene products also interact and produce an identifiable response. This is called the yeast two-hybrid system as it usually uses yeast genes as the detector marker. The disadvantage of both of these processes is that the experimenter must have an idea as to what the interacting proteins are before the experiment is performed. Mass spectrometric methods have recently been devised to overcome these shortfalls. In the first method the protein and interacting proteins are purified in the same way as the first non-MS method (i.e. immunoprecipitation and 1D SDS–PAGE separation). These are then excised from the gel and analysed, as previously described, by mass spectrometric methods in order to determine protein identification. Although this method can identify the interacting proteins it suffers from the fact that the immunoprecipitant is also present and can interfere with these detection processes. The second method uses mass spectrometry to identify the proteins in exactly the same fashion but purifies the protein of interest and its complexing proteins in a different manner. The purification of the interacting proteins (and also the quantitation of the binding) is performed by surface plasma resonance (SPR) in which the protein of interest is attached to a surface and a cell lysate or alternative sample is passed over the bound protein. Proteins that interact with this bound protein bind to it and thereby increase the amount of biological material on the SPR chip. This increase in biomaterial causes a change in the refractive index at the surface/solution interface. This change is detected by a change in the angle/wavelength that light is absorbed by the bound sample and the change can be related to the amount of material at this site. The complex can be eluted and studied by mass spectrometry as before. The advantage this process has is that it allows both the qualitative binding (which proteins bind) to be studied as well as the study of the quantitative binding (i.e. how much protein binds to the protein of interest (Nedelkov and Nelson, 2001).

A more recent development in the mass spectrometric study of protein–protein interactions has studied the peptides involved in such interactions (Mackun and Downard, 2003). The protein mixture is separated by native-PAGE in which the structure and properties of the proteins are not disturbed. Next either the proteins of interest or protein complexes are excised from the gel. If proteins are excised these are then washed with suspected complexing proteins (if complexes are excised this is not necessary). The complexes form are then subjected to limited proteolysis so that most the protein is digested but the peptides that form the complex remain bound to each other. The non-bound (and therefore non-complexing) peptides are then removed from the peptides that are bound together by molecular weight cut off filtration (usually approximately 3000 Da cut off). This leaves bound peptide complexes, these are then dissociated by acid treatment into individual peptides and analysed by mass spectrometry as described previously in order to determine which parts of the two proteins bind to form the complex.

6. Applications in plant proteomic studies

6.1. General

In strategic terms, the approaches with proteomic studies in plants are similar to those carried out in animal and microbial systems. Because of the sensitivity of the techniques involved in protein analysis and the fact that a relatively small perturbation can bring about marked changes in protein expression, care has to be taken to ensure that, apart from the experimental treatment or the particular perturbation involved, the tissues or cell systems subject to the comparative analysis are in precisely the same metabolic and developmental state.

The problem is highlighted by a study in the yeast *Saccharomyces cervisiae* in which perturbations in the galactose induction pathway, which consists of only nine central elements, resulted in changes in nearly 1000 yeast genes (Ideker et al., 2001). If this example is typical of the interconnectedness within biological systems, clearly even experimental perturbations that are intended to affect the activity of a single pathway or cellular element may have the potential for causing widespread changes in the quality and quantity of protein expression. Certainly, it has not been uncommon to observe changes in the abundance of seemingly unrelated, and frequently unidentified, proteins in response to an experimental treatment. Such data are difficult to explain in terms of the biology of a system and creates problems in the volume of data that has to be handled. The difficulties such problems create and a likely solution are discussed elsewhere (Aebersold, 2003).

Consideration also has to be given to the likely occurrence of post-translational modifications (discussed above). Techniques are now available to analyse the most common covalent modifications (Mann et al., 2001; Mann and Jensen, 2003; Zhu et al., 2003) but it is frequently a question of anticipating the occurrence of such modifications, so that the appropriate experimental strategy can be adopted at the outset of the investigations.

In addition to the complexity arising from post-translational modifications, the complexity is further increased when interactions between proteins and other molecules are considered. It is often the case that the

biological activity of a protein is not fully acquired until it is positioned in a particular cellular location, or it has become associated with other cellular components, including other proteins, nucleic acids, lipids or co-enzymes. Again, as part of the development of proteomics, techniques and strategies have been developed to investigate protein–protein interactions and even protein localization in the cell (von Mering et al., 2002; Zhu et al., 2003). Furthermore, developments in the form of technical advances in mass spectrometry, data handling and in systems biology, are in hand to deal with the complexity of biological systems (Aebersold, 2003; von Mering et al., 2002; Zhu et al., 2003).

6.2. Studies with organs and tissues

There have been many earlier studies in which levels of protein expression have been analysed with 2D PAGE. Most of these studies were designed to identify proteins that would serve as possible markers of different genotypes and phenotypes and to use in determination of phylogenetic relationships. However, most of them did not involve identification of the individual proteins, which was technically difficult before the development of MS and associated techniques for analysis of protein structure. This work is well summarised in the review by Thiellement et al. (1999). With the development of mass spectrometric and associated techniques for analysis of proteins work in this field has continued to make progress but is outside the scope of this review. However, excellent reviews of this work are available (Consoli et al., 2002; Thiellement et al., 2002; Zivy and de Vienne, 2000).

The 'typical' plant cell is likely to be made up of more than 20,000 different polypeptide species, and a maximum of only 3000 spots can be resolved and observed on a typical 2D PAGE gel (Lilley et al., 2001). Thus it is not realistic to imagine that analysis of a species proteome is readily achievable even with high-throughput facilities capable of analysing mixtures of polypeptides, notwithstanding the additional problems (discussed earlier) associated with analysis of post-translational modifications, low abundance etc. Not surprisingly then, there are few examples of attempts to do so. Where work is in progress, the approach has been to work with organ- or tissue-specific proteomes, in order to build a picture of the species proteome (Koller et al., 2002; Komatsu et al., 2003; Rakwal and Agrawal, 2003; Watson et al., 2003).

One study in rice illustrates the scale of the difficulties associated with such work (Komatsu et al., 2003). Analysis of the 2D SDS–PAGE gels of proteins from a range of rice tissues revealed a total of over 10,000 spots (not all of these would have been unique. Though no analysis was reported, it can be expected that many of the spots occurring on the individual gels would have represented the same protein expressed in the different tissues). Of these, 1220 were selected for analysis and 358 were identified from their sequences. While a figure of 30% for identification of the proteins in a selected set may be good, in terms of defining the whole proteome for a species, it is set against a background that the rice genome consists of about 30,000 genes, many of which may be expressed as proteins (Komatsu et al., 2003), and that draft sequences for two cultivars of rice are available (Goff et al., 2002; Yu et al., 2002). Genomic sequence data are an obvious aid to identification.

The most comprehensive analysis of a plant proteome was carried out in a separate study of rice, in which 2582 unique proteins from leaf, root and seed tissues were identified with a combination of 2D SDS–PAGE and LC-MS/MS, together with MudPIT (Koller et al., 2002). Despite the fact that the most abundant category for the identified proteins was that of unidentified function (ca. 33%), 20% of the proteins were classified as belonging to metabolic processes. A number of those belonged to so-called central metabolic pathways and were expressed in all three tissues. However, there was also evidence of tissue-specific expression for some of the proteins in this category.

Taking the same sort of organ-/tissue-specific approach, a survey of the proteomes of six tissues of the model plant barrel medic (*Medicago truncatula*) produced 2D SDS–PAGE reference maps from which 551 proteins were analysed using peptide mass fingerprinting and MALDI-TOF MS. (Watson et al., 2003). In this case the overall success rate for identification was 55%, a figure that is considered good in the absence of a sequenced genome, though the figure depended on the tissue in question. For example, identification was achieved for 43% of the proteins extracted from root tissue, while the figure for leaves was 76%, the difference presumably reflecting differences in the quality of the separations, and the information in the databases and availability of expressed sequence tags (ESTs).

This study also highlighted the difficulties that may arise when using only protein databases for identification. The overall success rate for identification using only the protein databases was 25%, while with the EST databases alone 40% of the total proteins were identified. Using both the protein and EST databases gave the success rate of 55% (the authors calculated the average overlap in the number of proteins identified in both databases as only 15%, indicating the complementarity of the databases but also showing the problems of relying solely on the protein databases for identification). Bearing in mind such results, there is an obvious need in any proteomic study to approach the process of assigning identity with caution.

Apart from the so-called 'descriptive' aspects of this study, were more functional features and amongst the wealth of data derived from this study are (a) 2D

SDS–PAGE reference maps, for future proteomic comparisons of genetic mutants, biotically and abiotically challenged plants, and/or environmentally challenged plants, (b) identification of proteins found only in a single tissue (39% of those proteins identified) which may therefore be used as physiological markers of tissue-specific protein expression and provide insight into the specialised physiological function of each tissue, (c) comparison between the levels of the identified proteins and the levels of the corresponding mRNAs, indicating that 50% of the proteins appear to be correlated with their corresponding mRNA, and (d) functional classification of the identified proteins.

There are few other reported studies on the same scale as the rice and medic projects; the majority of proteomic studies being concerned with measuring protein expression in tissues at different developmental stages (Salekdeh et al., 2002; Wilson et al., 2002), or in response to a particular environmental challenge (Chang et al., 2000; Shen et al., 2003a) or effector (Shen et al., 2003b). In some cases the focus has been a tissue but in attempts to reduce the complexity of the systems and the subsequent analysis, there are an increasing number of reports in which organelles, or a particular metabolic system has been the subject of investigation (Mo et al., 2003; Taylor et al., 2003). There are also reports in which the focus has been a particular protein, or set of proteins (Butt et al., 2001; Whitelegge, 2003; Xing et al., 2003). In general the approach has been one of determination of protein expression and any changes thereof in defined circumstances. The aim has been to relate any changes observed to a cellular mechanism. The problems associated with the number of genes that may change expression in response to a single 'trigger' and the subsequent difficulties with interpretation of the data have been referred to above. Such problems may be minimised, not only by focusing on a particular organelle or system but also by defining a clear hypothesis and the experimental design to address it.

Many proteomic studies have focused on changes in protein profile in response to environmental factors. For example, in a study to identify proteins associated with drought tolerance and drought avoidance in rice, analysis of 2D PAGE gels revealed a set of 42 proteins (from a total of 1000 observed on the gels) that changed in abundance in response to drought stress (Salekdeh et al., 2002). Out of this set of drought-responsive proteins 16 were identified by MALDI-MS or ESI-Q-TOF MS/MS. The majority of the identified proteins were from cellular pathways known to be responsive to drought, protein synthesis, photosynthesis and carbon metabolism, for example. However, four, an S-like RNAase homologue, actin depolymerising factor, rubisco activase (all up-regulated) and an isoflavone reductase-like protein (down-regulated), had not previously been associated with the drought response. The significance of the changes in regulation of these particular proteins remains unknown and so illustrates some of the difficulties with the descriptive approach to proteomic studies. A great deal of data may be generated from proteomic studies but in the absence of additional physiological or biochemical experimentation, there may then be great difficulty in interpreting the data in biological terms.

A similar descriptive style of approach was taken with a study of senescence in white clover (*Trifolium repens* (L)), the aim of which was to identify protein changes that might be involved in the induction and progression of senescence (Wilson et al., 2002). In this case, 590 spots were observed in 2D SDS–PAGE gels, 40% of which showed changes in abundance related to senescence (some increasing and some decreasing). Twelve of those occurring in a chloroplast fraction were examined by MALDI-TOF MS, although, because of the lack of sequence data for *T. repens*, only six were identified. Perhaps not surprisingly, given the organelle involved, the six included rubisco large and small subunits, a rubisco activase and the 33 kDa protein of the photosystem II oxygen-evolving complex. All these declined during senescence. Given the number of proteins that changed in abundance during senescence and that the authors have been able to select only a small fraction for identification, there is much further work to do before their aim can be reached. A novel aspect of this study was identification of the coat protein of the white clover mosaic virus among the proteins present in the extract, giving rise to the concept of identifying viral infections in plants with an MS approach. Indeed, a recent report documents identification of the coat protein from Tobacco Mosaic Virus in protein extracts from infected tobacco plants using LC-MS/MS (Cooper et al., 2003). The technique was also capable of identifying a virus, the identity of which was unknown to the investigators, though in neither case was the data sufficient to allow identification of the virus strain.

One of the difficulties facing the investigator is where to focus attention for further analysis when changes in abundance are observed for a large number of proteins. That difficulty was partially overcome in a study of protein synthesis during hypoxic acclimation to anoxia of *Zea mays* seedlings, by the use of [^{35}S]methionine. The methionine was used to label proteins synthesised during the acclimation period, thereby focusing on proteins likely to be involved in the acclimation process (Chang et al., 2000). Analysis of fluorographs from 2D SDS–PAGE gels revealed ca. 260 proteins synthesised during the acclimation period. Those proteins were synthesised in both tissue under the hypoxic regime and the normoxic controls. However, synthesis of ca. 10% of the proteins was enhanced in the hypoxic tissues, suggesting a role in the acclimation process. Forty-eight of those proteins were selected for analysis by delayed

extraction MALDI-TOF, including both those whose rate of synthesis was enhanced and those whose rate was depressed during acclimation. Forty-six were identified, a remarkably high success rate. Predictably, among the identified proteins were enzymes associated with anaerobic metabolism, for example alcohol dehydrogenase, but there were also changes observed in the expression of proteins associated with protein turnover and targeting. The use of [^{35}S]methionine enabled attention to be focused on a specific subset of proteins. However, when it came to identification, the investigators were still obliged to take into account the resolution achievable by 2D SDS–PAGE when selecting proteins for analysis by MS. Nevertheless, the study has given some insight to the changes occurring during acclimation to anoxia and provides focus for future investigations.

In order to reduce complexity compared to the whole plant, a study of wound-response related proteins in rice was carried out in the leaf sheath (Shen et al., 2003a). Nevertheless, over 400 protein spots were detected in the 2D SDS–PAGE gels. In an attempt to focus on those related to wounding, image analysis was carried out on 2D gels to identify proteins whose expression altered during the response. Twenty-nine proteins were observed to change in abundance (19 of which were down-regulated). Of these 14 were analysed for structure, 11 by MALDI-TOF MS. Again, the selection of proteins for analysis was directed more by the amount of protein in the gel and the resolution achieved, rather than critical experimental evidence. Even so, four of the proteins were among recognised wound-response proteins (Bowman-Birk trypsin inhibitor, receptor-like protein kinase (2), calmodulin-related protein), and several signal-transduction-, photosynthesis- and stress-related proteins were among those that changed expression in relation to wounding. Several of these had been identified in previous studies (for example, Rakwal and Komatsu, 2000). Interestingly, in the case of the receptor-like protein kinases and of calreticulin, two forms of the protein were identified, each with the same M_r but differing in pI, suggesting to the authors that the proteins have been subjected to post-translational modification, though no attempt was made to examine this aspect further.

Calreticulin is a Ca^{2+} binding protein, which can also act as a molecular chaperone (Crofts and Denecke, 1998). Changes in its expression pattern were also observed in a study to characterise proteins responsive to gibberellin (GA) in rice leaf sheath (Shen et al., 2003b). Calreticulin was one of 32 proteins that analysis of 2D SDS–PAGE gels showed changed in abundance in response to treatment with GA (21 of which were up-regulated), all of which were identified by sequencing or delayed extraction MALDI-TOF MS. In this case the authors were able to demonstrate that in response to treatment with GA the abundance of one isoform of calreticulin was decreased while the other increased, suggesting the conversion of one form to another by post-translational modification. Furthermore, over-expression of calreticulin inhibited growth in callus and seedlings, though the effects on GA on the transgenic plants was not assessed. In this case the proteomic approach was supplemented with a physiological/biochemical experimentation, which added considerably to interpretation of the data obtained.

Some reduction of complexity is evident in two reports, which although descriptive in approach provide interesting insights into two related biological systems. In a study to examine the distribution of proteins present in the phloem of different source and sink organs, MALDI-TOF MS was used to identify the patterns of protein in phloem exudates from different tissues (seedling, flowers and leaf tissue as well as large and small fruits) (Kehr et al., 1999). Because the range of proteins present in the exudates was limited, the authors were able to carry out MALDI-TOF MS without the need for preceding purification, to identify molecular masses, and show that some proteins were present in whatever the source of the exudates. However, conversely, some were specific to the tissue concerned. Identity of some of the proteins was established using tryptic digestion and MALDI-TOF MS following 1D SDS–PAGE on the phloem exudates, and indicated that post-translational modification occurred to one protein (PP2).

Some interesting data were provided by an examination of apoplastic extracts from three species, *Arabidopsis thaliana*, *Triticum aestivum* and *Oryza sativa* (Haslam et al., 2003). 2D SDS–PAGE was used to separate the proteins in the exudates prior to identification of selected proteins by MALDI-TOF. The principle components identified reflected the functions of the cell wall and apoplastic space and included proteins involved in defence, germin-like proteins and glucanases, and those involved in cell expansion, β-D-glucan glucohydrolase. Comparison of the profile from *A. thaliana* with that from *O. sativum* showed what were interpreted as species differences. However, the majority of proteins could not be identified, indicating yet unknown functions of the apoplast.

6.3. Studies with seeds

A number of proteomic analyses have been carried out with seeds; many of them, because of the significant commercial interest, with wheat (*Triticum aestivum* L.). In an early study, analysis was restricted to comparison of the patterns of spots on 2D gels (Skylas et al., 2001), with the intention of developing methods to discriminate between cultivars. In a subsequent study of the genetic control of wheat proteins MALDI-TOF MS was used in the identification of two novel subunits of high molecular mass glutenin (Islam et al., 2002). Other studies

have revealed diversity amongst the puroindoline content (lipid-binding proteins that affect the properties of bread doughs and may be associated with grain softness) of wheat varieties (Branlard et al., 2003), and have identified changes in the expression of wheat proteins resulting from high temperatures during grain filling, which again may influence dough quality (Majoul et al., 2003).

The most comprehensive MS-based characterisation of the proteins expressed in wheat seed was carried out with amyloplasts and amyloplast membranes, in a study to characterise protein expression in relation to control of starch synthesis and grain-filling (Andon et al., 2002). Identification was made in the case of 108 of the ca. 200 proteins observed on 2D gels from whole amyloplasts and in the case of amyloplast membranes 63 proteins were identified. The study provided interesting data on the occurrence and distribution of starch storage and synthesis related proteins, and on transport proteins. Another interesting aspect of this study is the approach taken to identify proteins. Because little genomic sequence data were available for wheat, the authors used HPLC followed by tandem MS of individual peptides rather than peptide mass fingerprinting for protein identification. Not only did this influence the speed and accuracy of the analysis but it allowed identification of the proteins from the amyloplast membrane fractions with 1D electrophoresis alone. The lack of genomic sequence data meant that protein identifications were derived from protein sequences for cereal crops other than wheat. In fact, 46% of the proteins were identified from peptides present in a propriety rice genomic database (Syngenta).

A particular aspect of wheat seed metabolism was examined in a study in which protein targets of a NADPH/thioredoxin/thioredoxin reductase system were identified with the thiol-specific fluorescent probe monobromobimane (Wong et al., 2003). The probe labelled only those proteins reduced by the thioredoxin/thioredoxin reductase pair. As a consequence, that particular subset of proteins was readily located on the subsequent 2D SDS–PAGE gel and became the focus of further analysis, including enzyme assays. Labelling with the probe also meant that the authors could identify the protein targets solely by comparison to an extensive 2D map of the soluble proteins from wheat endosperm, which had been constructed previously using MS techniques. The combination of approaches used in this study led to identification of protein targets of thioredoxin that are involved in starch and protein breakdown, and oxidative stress. Furthermore, the results indicated the importance of thioredoxin in control of enzyme activity and redox balance in the life cycle of the seed.

As may be expected, some studies have been carried out with *Arabidopsis* seeds, to take advantage of the availability of the complete genome sequence. In one, protein profiles were established with 2D SDS–PAGE gels of protein extracts from seeds at different stages of germination and subjected to different pre-treatments (Gallardo et al., 2001). Of the total (1300) seed proteins resolved in the 2D gels 74 changed in abundance but only 67 were identified by MALDI-TOF MS. Some had been associated with germination in previous studies, though the finding that an actin isoform and a WD-40 repeat protein are involved in the imbibition process was novel. These studies were subsequently extended with the use of a GA-deficient mutant, and treatment with paclobutrazol, an inhibitor of GA synthesis (Gallardo et al., 2002). This approach enabled identification of several proteins the abundance of which is controlled by GA, including α-2,4 tubulin (a component of the cytoskeleton), S-adenosyl-methionine synthetase and β-glucosidase, suggesting that GA controls protein expression at different stages of the germination process. A more recent study has focused on changes in protein expression at specific stages of seed filling in *M. truncatula* (Gallardo et al., 2003). Changes in abundance of 120 proteins were detected and 84 of them were identified through peptide mass fingerprinting by MALDI-TOF MS. Changes in some of these were associated with cell division (β-tubulin and annexin), or protein storage (vicilins, legumins and convicilins), or carbon metabolism (sucrose synthase, starch synthase), or cell expansion (actin), depending on the particular stage between highly active to quiescent state at which the expression profiles were observed. In a study of changes in protein expression during germination, lipid transfer proteins that may play a part in recycling of lipids during germination of seeds of *Euphorbia lagascae* were identified by 2D SDS–PAGE (Eklund and Edqvist, 2003).

6.4. Organelles

The complexity of analysing proteomes is reduced somewhat by focusing on the proteomes of subcellular organelles. Such organelles represent distinct functional units and in many cases their activities have been well defined. Interpretation of proteomic data can therefore be carried out against a functional background. The advantages, and problems, mainly issues concerning preparation, purity and correct identification of organelles, are discussed by Mo et al. (2003) and Taylor et al. (2003). Given the availability of the *Arabidopsis* genome there is a focus on the use of this tissue as a source of organelles (see for example, Prime et al., 2000).

6.5. Chloroplast

It is not surprising that the chloroplast has been the subject of a number of proteomic studies, some of which are described in earlier reviews (Taylor et al., 2003;

van Wijk, 2000). Much is known about the biochemistry/physiology of the chloroplast and the genomes that encode its proteins and control expression of them. However, despite the reduction in complexity of its proteome compared to the whole cell, it should be remembered that estimates of the number of proteins located in the chloroplast range between 2100 and 3600 (Leister, 2003). Furthermore, the great majority of these are encoded by the nuclear genome and have to be transported there, the chloroplast genome containing only some 60–200 open reading frames (Leister, 2003).

From a functional and structural point of view, the chloroplast is composed of several compartments, each with its own subset of proteins. Because of differences in character, each subset requires different experimental strategies for study of its proteome (van Wijk, 2000). The chloroplast envelope is the site of many functions, including metabolite transport, glycerolipid synthesis and export of fatty acids (Joyard et al., 1998). The difficulties involved in studying this subproteome include purification of the membranes, the hydrophobic nature of its proteins, the resultant problems in extraction and 2D electrophoresis, and the low abundance of some of its proteins (Froelich et al., 2003; van Wijk, 2000). To overcome the hydrophobicity of the membrane proteins, methods involving extraction with combinations of organic solvents, high salt concentrations and detergents have been developed (for example, Ferro et al., 2002, 2003; Seigneurin-Berny et al., 1999). Using 1D SDS–PAGE and tandem MS for analysis of proteins extracted from purified chloroplast envelope membranes from spinach, 54 proteins were identified, 27 of which had not previously been identified in the envelope (Ferro et al., 2002). Most of them had multiple α-helical transmembrane regions and were characterized as likely to be transporters. By comparing features between known and the newly identified transporters, the authors constructed a virtual plastid envelope integral protein database from a search of the complete *Arabidopsis* genome database, and identified further candidates for transport functions in the envelope. Using a similar data-mining approach, computational methods alone were used to predict chloroplast envelope proteins from the *Arabidopsis* nuclear genome (Koo and Ohirogge, 2002). In this case, 541 candidate proteins were identified and 183 of these were assigned a putative function. A similar 'in silico' search of the *Arabidopsis* genome, for genes encoding proteins with predicted β-barrel structures and signals that would direct them to the chloroplast, led to identification of several candidates (Schleiff et al., 2003). Such exercises may be useful in alerting investigators to the potential presence of low abundance proteins that may ordinarily be missed in an experimental investigation.

Arabidopsis envelope membranes were also the subject of a study in which different extraction methods were used to retrieve proteins ranging from the least to the most hydrophobic (Ferro et al., 2003). In this case LC-MS/MS led to identification of ca. 100 proteins, most of which were known to be located in the envelope. The functions associated with this subset of proteins included ion and metabolite transport, protein import and chloroplast lipid metabolism, though nearly 30% had no known function. Some soluble proteins, including proteases, were also associated with the envelope membranes. Interestingly, a 350-kDa ClpP protease complex, consisting of 10 different subunits, was identified earlier in *Arabidopsis* chloroplasts though this one is associated with the thylakoid membranes (Peltier et al., 2001). The use in this study of Blue-Native (BN) PAGE in the first dimension left the protein–protein interactions intact and allowed for separation of the complex as a whole. Using a combination of MALDI-TOF MS and nanoelectrospray MS/MS, the subunits were characterized to the point where several truncations and errors in intron and exon prediction of the annotated Clp genes were corrected.

As an alternative to 2D electrophoresis because of problems associated with its use with the highly hydrophobic proteins of the envelope, one recent study of the *Arabidopsis* chloroplast membrane involved two different strategies, off-line MudPIT, and 1D gel analysis combined with proteolytic digestion and LC-MS/MS (Froelich et al., 2003). The authors also made use of novel methods for the isolation of chloroplasts and chloroplast membranes. In all, the study led to the identification of 392 non-redundant proteins (149 exclusively from off-line MudPIT, 109 exclusively from LC-MS/MS), the majority of which, surprisingly, did not contain a transmembrane domain. Although ca. 60% of the proteins could not be assigned a definite function, functional categories that were recognised included, protein import, metabolite translocators, fatty acid metabolism, oxylipin metabolism, saccahride transporters and ABC transporters.

Import into the chloroplast was the subject of another study making use of chloroplasts from *Arabidopsis*. A component of the translocon complex responsible for import of proteins into the chloroplast, Toc34, encoded by two homologous genes *atTOC33* and *atTOC34*, was examined for its role in protein import (Kubis et al., 2003). Focus was given to this analysis by the use of an *atTOC33* knockout mutant combined with difference gel electrophoresis (DIGE), in which proteins from the mutant and wild-type chloroplast were labelled with complementary CyDyeDIGE flours. The combination allowed recognition and identification (by LC-MS/MS) of proteins that differed in abundance in the mutant and wild-type, which in turn indicated that atToc34 is involved preferentially with the import of photosynthetic proteins.

Some caution must be used when considering targeting of proteins to intracellular sites however. In a

study that made use of intact protein mass measurement (see below), a dataset of 58 nuclear-coded thylakoid integral membrane proteins was created, to examine their transit peptide cleavage sites (Gomez et al., 2003). The dataset was used to challenge Web tools for predicting organelle targeting and transit peptide proteolysis sites The results indicate that whereas many of the algorithms were capable of correctly predicting the target site, none of them adequately predicted the transit peptide processing site of the membrane-spanning proteins of the thylakoid.

Problems with hydrophobicity are also evident with analyses of the integral proteins of thylakoid membranes. However, techniques to deal with such problems and allow ESI-MS of integral membrane proteins have been developed (Whitelegge et al., 1998, 2002). The techniques developed, which included separation of individual proteins with a combination of size exclusion and reverse phase chromatography, provide extremely accurate mass measurements of intact proteins (intact mass tags [IMTs]) and by using ESI on a quadrapole mass spectrometer routinely achieve 0.01% mass accuracy (Whitelegge et al., 1998). A mass spectrum of an intact protein defines the native covalent state of the product of a gene, and so reveals post-translational modifications and any resultant heterogeneity (Whitelegge et al., 1998). For example, a resolution of 0.01% (100 ppm) would allow detection of two protein species differing by a single methionine oxidation (+16 kDa) up to at least 50 kDa M_r (Whitelegge, 2003). Similarly, the accuracy of the measurement of IMTs provided by ESI can lead to unequivocal assignment of a protein to its respective gene. Consequently, measurement of IMTs is an important step in the analysis of a protein, though not all reported work is carried out to a 100 ppm standard.

In one study, size exclusion and reverse phase chromatography was used to separate the protein subunits of the cytochrome b_6f complex from spinach before they were analysed by ESI-MS (Whitelegge et al., 2002). In this case, the combination of HPLC and ESI-MS gave an accurate description of the molecular structure of an integral membrane complex. This report is significant because, with its use of size exclusion and reverse phase chromatography combined with ESI-MS operating to 0.01% mass accuracy, it is the first to provide full subunit coverage of a protein complex. It also illustrates the advantages of measuring ITMs, since by this means cytochrome *b* was observed to be associated with a haeme residue, though the two haemes known to be associated with cytochromes *b* are non-covalently attached and should not have remained associated during the ESI process. The subsequent crystal structures confirmed the presence of an unknown haeme covalently linked to cytochrome *b*, first revealed by the accuracy of the mass measurement (Stroebel et al., 2003). Furthermore, the low M_r subunits of the cytochrome b_6f complex were sequenced without prior cleavage using MS/MS, and for one of them the structure differed from the translation of the spinach chloroplast genome, indicating a DNA sequencing error or an undiscovered RNA editing event.

The techniques are also capable of working with complex mixtures and intact mass measurements have been used to define the chloroplast grana proteome in respect of some 50–100 proteins (Gomez et al., 2002; Schwartz et al., 2002). In the case of the study with pea and spinach, approximately 90 IMTs were detected, corresponding to approximately 40 gene products with variable post-translational covalent modifications; based upon coincidence between measured mass and that calculated from genomic sequence, 30 of these were provisionally identified (Gomez et al., 2002). In the same study, analysis of isolated photosystem II complexes allowed detection and resolution of a minor population of D1 (PsbA) that was apparently palmitoylated. Furthermore, the study provided evidence of a second phosphorylation site on PsbH and a new phosphoprotein was proposed to be the product of *PsbT*. This study also illustrates the need to sometimes work with highly purified sub-fractions when dealing with complex mixtures to avoid bias against low abundance and membrane proteins. ESI-MS has also been used in other studies to analyse photosystem II reaction-centre subcomplexes (Sharma et al., 1997; Zheleva et al., 1998). In a later study, LC-ESI-MS was used successfully to characterise the light-harvesting proteins of photosystem I from four monocot- and five dicot species (Zolla et al., 2002). The accuracy with which the intact molecular masses were measured enabled unequivocal assignment of the proteins, while differences between measured mass values and those calculated from nucleotide-derived amino acid sequences again suggested errors in DNA sequence determination.

A systematic analysis of the lumenal and peripheral proteins from the thylakoids from pea chloroplasts showed the presence of at least 200 different proteins, even after isoforms and post-translational modifications had been eliminated (Peltier et al., 2000). Sixty-one of the proteins were identified with a combination of MALDI-TOF-MS and ESI-MS/MS, and Edman sequencing, 31 of which were assigned a function. Subsequently, the same group has identified a number of lumenal and peripheral proteins from the thylakoids from *Arabidopsis* chloroplasts using ESI-MS/MS and peptide mass fingerprinting, and have estimated the lumenal proteome to consist of 71 proteins (Peltier et al., 2002).

In a study that gives an interesting insight to the photosynthetic light reactions, MALDI-TOF-MS and ESI-MS was used to identify proteins that were phosphorylated in photosystem II, including the D1, D2, and

CP43 core proteins and the mature light-harvesting polypeptides LCHII (Vener et al., 2001). Not only were the specific sites of phosphorylation identified but it was also possible to assess the effect of light and heat on the degree of phosphorylation. In a study that made use of mutated thioredoxin as the immobilised ligand in an affinity chromatography step, a number of protein targets of thioredoxin were identified in the chloroplast with LC-MS/MS (Balmer et al., 2003). These included enzymes from pathways previously unrecognised as thioredoxin linked, for example plastid-nucleus signalling. Detailed characterization of the proteins present in the ribosomal 30 S and 50 S subunits of spinach chloroplasts, using a variety of proteomic techniques including ESI-MS and LC-ESI-MS has identified post-translational modifications to a number of the components (Yamaguchi and Subramanian, 2000).

A recently developed approach, termed top-down Fourier Transform Mass Spectrometry (FTMS), has been shown to be of great potential by its application to a study of chloroplast proteins in *Arabidopsis* (Zabrouskov et al., 2003). Of 3000 proteins predicted by the genome sequence, 97 proteins had been identified by digestion of the purified proteins and matching the MS-determined mass of the resultant peptides against those expected from the DNA-predicted proteins. By top-down FTMS, in which the molecular ions from a protein mixture had their masses determined to ± 1 Da and were then fragmented by FTMS, of seven selected proteins in which the amino acid sequence was completely determined, six differed from the genome-predicted sequence. This top-down FTMS approach is thus very effective for determining cleavage sites of signal peptides and locating post-translational modifications, in addition to differentiating highly similar proteins.

6.6. Mitochondrion

In one of the earlier studies to make use of MS analysis, the proteins from *Arabidopsis* mitochondria were examined (Kruft et al., 2001). Only 52 proteins from a total of 800 were identified, using a combination of immunoblotting, direct sequencing and MS. A greater success rate was achieved for the soluble proteins from the mitochondria from pea leaves, roots and seeds (Bardel et al., 2002). The 2D gel maps from green chloroplasts contained over 400 spots, ca. 70% of which were identified with a combination of Edman degradation, MALDI-TOF-MS and ESI-MS/MS. A number of the polypeptides occurred as multiforms, indicating isoforms or post-translational modifications. Furthermore, some proteins were specific to root or seed mitochondria, indicating the influence of tissue differentiation on the mitochondrial proteome.

An extensive examination of the proteins from rice (*Oryza saliva*) mitochondria has been carried out with a combination of 2D SDS–PAGE, BN-PAGE (Blue Native PAGE) and LC-MS (Heazlewood et al., 2003a). The use of BN-PAGE, in which the native proteins are complexed with the Coomassie dye before electrophoresis, allowed characterisation of the electron transport chain complexes. Combined with LC-MS/MS on a digest of the total protein preparation and LC-MS of the separated proteins, 136 mitochondrial proteins were identified. BN-PAGE was also used to fractionate the protein complexes of the *Arabidopsis* mitochondrion, prior to separation of the components from each complex by SDS–PAGE (Giege et al., 2003). Subsequent MS analysis led to the identification of 29 proteins of the complexes, including chaperones, transporters, proteins of the respiratory chain and proteins novel to plant mitochondria.

BN-PAGE again featured in an analysis of supercomplexes from the *Arabidopsis* mitochondrion (Eubel et al., 2003). Various concentrations of non-ionic detergents were used to treat mitochondrial preparations from *Arabidopsis*, *Solanum tuberosum*, *Phaseolus vulgaris* and *Hordeum vulgare*, and the resultant protein complexes were separated by BN-PAGE and 2D gel electrophoresis. Three supercomplexes were recognised on BN-PAGE gels. The composition of each complex was determined by MS analysis, showing that the supercomplexes were formed by combinations of the conventionally recognized mitochondrial complexes (I–IV), with the source species having some influence on supercomplex composition.

In subsequent investigations, BN-PAGE separation of *Arabidopsis* mitochondrial complexes was used to examine the F_1F_0 portion of the ATP synthase (Heazlewood et al., 2003b). MS analysis was used to identify the five proteins of the F_1 complex, while four proteins of the F_0 complex were identified, two of which matched to the mitochondrial encoded *orfB* and *orf25* products. The *orfB* encoded product was also identified among a series of divalent metal ion-binding proteins present in the *Arabidopsis* mitochondrial proteome (Herald et al., 2003). A set of metal binding proteins was identified through their mobility shifts in the presence of divalent cations during 2D diagonal SDS–PAGE. Subsequent MS/MS analysis led to identification of known divalent cation-binding proteins and other proteins not previously known to bind metal ions, including some of unknown function.

Mitochondrial carrier proteins are responsible for maintaining communication with the cytosol, though in previous studies of mitochondrial proteome analyses none were identified (Kruft et al., 2001; Millar et al., 2001). Subsequently, using a modified method for the extraction of hydrophobic proteins, followed by 1D electrophoresis and MS/MS-based sequencing of doubly charged peptides, Millar and Heazlewood (2003) identified six carrier proteins from the mitochondrion of *Arabidopsis*. There has also been a report of the impact of oxidative stress on proteins of *Arabidopsis*

mitochondria (Sweetlove et al., 2002). In this case proteins that changed in abundance in response to the stress were analysed by 2D SDS–PAGE followed by MS/MS. Several, including subunits of ATP synthase and complex I, decreased, while components involved in antioxidant defence, thioredoxin-dependent peroxidase for example, increased.

6.7. Other organelles

The cell wall represent a major investment for the plant cell and the proteins of the cell wall from *Arabidopsis* have been studied, with a view to understanding the role of the wall in development (Chivas et al., 2002). In this study, proteins sequentially extracted from the wall were analysed by 2D SDS–PAGE and MALDI-TOF MS. Not only were classical cell wall proteins identified in the extracts but proteins of unknown function and proteins normally associated with other cellular compartments were also present. In a subsequent study, sequential extraction procedures were again used to prepare proteins from *Arabidopsis* cell walls, and MS was used to identify known and previously unknown proteins of the wall (Borederies et al., 2003).

The proteome of the nucleus of *Arabidopsis* was analysed using 2D SDS–PAGE and MALDI-TOF MS, leading to identification of 158 different proteins, 54 of which changed in abundance in response to cold stress (Bae et al., 2003). In a separate study, the nuclear matrix of *Arabidopsis* was examined with a combination of approaches, including confocal and electron microscopy, 2D SDS–PAGE, and 1D SDS–PAGE followed by ESI-MS/MS (Calikowski et al., 2003). Approximately 300 spots were resolved by 2D electrophoresis, 36 of which were identified by MS, including recognised nuclear proteins, ribosomal components and proteins of unknown function.

In developing oil seeds the endoplasmic reticulum (ER) plays a central role in synthesis, sorting and storage of protein-lipid reserves, while in germinating seeds degradation predominates. These aspects of ER metabolism were examined in a study of ER from developing and germinating seeds of *Ricinus communis* (castor), which involved 2D SDS–PAGE, and 1D SDS–PAGE combined with MALDI-TOF MS mass fingerprinting (Maltman et al., 2002). A number of proteins were identified, including calreticulin oleate-12-hydroxylase, involved in ricinoleic acid synthesis, and differences in protein expression profiles of the developing and germinating seeds were observed.

6.8. Plant proteins with specific properties

A number of studies have been directed towards further characterisation of specific molecules and/or their function. For example, 30 glycosylphosphatidylinositol (GPI)-anchored proteins (GAPs) were identified in *Arabidopsis* callus cells, using a combination of 2D DIGE and 1D SDS–PAGE with LC-MS/MS (Borner et al., 2003). The analysis revealed the occurrence of GPI anchoring in several families of proteins, including glucanases, phytocyanins and receptor-like proteins, increasing our understanding of polarized targeting and related cell surface processes.

Given its significance in regulation of cell processes it is not surprising that phosphorylation has been the object of proteomic studies. In *Arabidopsis* suspension-cultured cells a number of proteins were phosphorylated in response to treatment with fragments of chitin or with flagellin, a fungal and a bacterial elicitor of defence responses (Peck et al., 2001). 2D SDS–PAGE combined with MS was used to identify one protein, AtPhos43 the phosphorylation of which was dependent on the activity of a receptor-like kinase. Phosphorylation, via a mitogen-activated protein kinase (MAPK) pathway, was also shown by MS methods to be involved in the induction of β-1,3-glucanase- and endochitinase activities in tomato, in response to *Pseudomonas syringae* (Xing et al., 2003). In an effort to improve detection of phosphoproteins and identification of their phosphorylation sites, phosphoproteins from *Arabidopsis* plasma membranes were digested with trypsin and the digest subject to anion exchange chromatography followed by IMAC before identification of the phosphopeptides by LC-MS/MS (Nuhse et al., 2003). The technique was successfully used to identify two previously unknown phosphorylation sites on an H^+-ATPase. In a study of mitochondria from *Solanum tuberosum* 14 proteins that had not been previously recognized as phosphorylated were identified, including enzymes of the tricarboxylic acid cycle (Bykova et al., 2003). Proteomics was used in a study of aquaporin isoforms in the plasma membrane from *Arabidopsis* roots (Santoni et al., 2003). At least five different aquaporin isoforms were identified and there was evidence for different degrees of phosphorylation of some of the isoforms.

Proteomic analysis has also been used to identify kinase activity in plants. For example, a glyoxysomal protein kinase was identified in the glyoxysomes from cotyledons of *Arabidopsis*, its amino acid sequence and protease sensitivity indicating it is a peripheral membrane protein with the kinase domain inside the glyoxysome (Fukao et al., 2003). A complement of protein tyrosine kinases has been identified in *Arabidopsis* through bioinformatics screening of the *Arabidopsis* proteome, and a number of *Arabidopsis* proteins phosphorylated on tyrosine residues were detected (Carpi et al., 2002).

Nitric oxide (NO) is now recognised to play a significant messenger role in plants, though there is still controversy concerning its endogenous source. No ob-

vious homologue of mammalian-like nitric oxide synthase (NOS) has been found in the *Arabidopsis* genome, though the presence of NOS in plants has been inferred from identification of proteins that react positively with mammalian NOS antibodies. In a recent study, 15 out of 20 proteins that interacted with antibodies to mammalian NOS were identified (Butt et al., 2001). None of them had functions related to NO metabolism, indicating the difficulties associated with identification of proteins through immunoreactivity alone.

7. Future directions

Various modifications in protocols and new procedures have recently been incorporated into Proteomic studies. An aid to the examination of low abundance proteins has been the incorporation of narrow range pH gradient strips, enabling the detection of as little as 300 molecules of a protein (Hoving et al., 2000). Various means of pre-electrophoretic fractionation have been beneficially employed including anion exchange chromatography (Butt et al., 2001), sucrose density gradient centrifugation (Hanson et al., 2001), hydrophobic interaction chromatography (Stevanovic and Bohley, 2001), and the use of anti-phosphotyrosine antibodies (Stancato and Petricoin, 2001). Graphite has been utilized as an alternative means of desalting and concentrating prior to MALDI MS (Larsen et al., 2002). Other notable developments have included laser dissection microcapture (Xu and Caprioli, 2002), direct tissue analysis (Schwartz et al., 2003) and the application of surface plasmon resonance, already used in concert with MS for small molecules (Sonksen et al., 2001) to select proteins.

A commercial survey of users and suppliers within the proteomics field (Cambridge HealthTech Institute, 2003) gives insight into the current emphases and future direction of proteomic studies. The followers of each see strong advantages and disadvantages in the use of the competing technologies of electrophoresis and chromatography, with a strong growth in the use of ultrafiltration prior to the use of both. Protein antibody chips, LC/MS and MS/MS are seen as key to the future development of proteomics at most sites.

Perhaps surprisingly 1D gels are still used more frequently than 2D, with SDS–PAGE the most common matrix in both modes, and robotics used only in the minority of these uses. The majority of such gels are used for image analysis, with increased automation and improved resolution and sensitivity cited as the key improvements in this area. A large increase in the use of LC separations is forecast, with paradoxically resolution specified as both a major advantage and disadvantage of LC use. Reverse phase and affinity LC are currently the most used LC systems, with the most common five LC setups accounting for half of proteomics-related chromatography. The use of capillary electrophoresis as a separatory technique is also expected to expand.

The most commonly utilized MS techniques in proteomics are currently MALDI-TOF and electrospray-QTOF MS/MS, with increased use of electrospray ion trap MS, with its capacity of MS^n analyses, and MALDI-TOF/TOF, with the capacity for MS/MS analysis of intact proteins as well as digests, anticipated. Phosphorylation remains, and appears likely to remain in the immediate future, the most frequently studied post-translational modification. Mascot and SwissProt are the most frequently accessed peptide mass and protein sequence databases respectively. An interesting recent proposal concerning proteomics databases is that of Taylor et al. (2003), suggesting a structured approach including a standard representation of both the methods used and data generated in a proteomics repository known as PEDRo (Proteomics Experiment Data Repository).

In our view much of the direction and eventual success of proteomic research will depend upon the technological developments of mass spectrometers and associated instruments. There are many such developments both current and imminent. For example a softer version of MALDI, where less fragmentation occurs, atmospheric MALDI (APMALDI), has continued to be developed (Laiko et al., 2000; Doroshenko et al., 2002; Baldwin et al., 2001) and is relatively simple to interface to mass analyzers. Surface enhanced laser desorption ionisation (SELDI) has been shown to be very powerful for selective ionization of protein fractions (Chapman, 2002). Signal amplification using ''spot on a chip'' (Ekström et al., 2001) has been reported with a signal amplification of up to a factor of 50, compared to ordinary sample preparation. Advanced preparation and presentation of the MALDI target plate is an area for further improvement. There are many exciting reports of protein chips used to immobilise proteins prior to detection (Lee and Mrksich, 2002).

Interfacing of MALDI to separation techniques is ongoing (Rejtar et al., 2002; Preislar et al., 2002). Capillary electrophoresis (CE) has received considerable attention due to its potential for protein/peptide analysis and high sensitivity. Deposition on to a continuous Mylar tape has shown separation efficiencies of 400,000 theoretical plates and low attomol sensitivity (Preislar et al., 2002). More recently chromatographic coupling with ''snail trail'' deposition onto a MALDI target has demonstrated the potential of off-line coupling of high resolution separations to MALDI-MS and MALDI-MS/MS using vacuum deposition for the analysis of complex peptide mixtures from protein digests (Rejtar et al., 2002). For MALDI analysis of high-resolution chromatographic separations, high repetition lasers are now being employed. A common chromatographic

method for achieving high sensitivity for fast LC separations has been peak parking (Davis and Lee, 1997); whilst Eksignet (2003) has demonstrated precise control of LC flow rate in the low nanolitre (nl) range which may supercede the ''peak parking'' technique.

Accurate mass measurement has been a powerful technique throughout the history of mass spectrometric development. Classically it has been used for elemental composition identification, for masses less than ~800 u. Accurate mass measurement has been applied to proteomic analysis and can improve the specificity of the technique and when combined with database searching provides improved identification. Recent developments have been made with internal mass calibration using alternate reference/analyte sprayers (Wolff et al., 2001; Williams et al., 2003). Additionally, in 2003 mass spectrometer manufacturers designed time-of-flight instruments with mass measurement accuracies of 3 ppm. One issue with time-of-flight has been the drift of mass measurement over time. These problems are now being addressed by the instrument companies. FTICR has continued to be the premier accurate mass measurement instrument with a routine capability of 1–2 ppm accuracy. An comparative study of accurate mass measurement, primarily on small molecules, has been made by the Laboratory of the Government Chemist (LGC Ltd, Middlesex, UK) which surveyed participating laboratories in UK, Europe and USA. The relative merits of the most commonly used mass analyzers were examined for a wide range of users (Bristow and Webb, 2003). The report captures the relative capabilities of the current range of mass spectrometers and the protocols used to record accurate mass measurements of small molecules. With peptides an approach based on accurate mass measurements extends the Mudpit technology described earlier: it takes advantage of the outstanding mass measurement capability (<1 ppm full scan) of FTMS and is applied to allow high-throughput proteome measurements since only the distinctive accurate masses and LC elution time are needed (Lipton et al., 2002).

In principle TOF is the most sensitive mass analyzer and designers have for many years attempted to achieve high performance ''TOF–TOF'' MS/MS analyzers. The first commercial high performance TOF–TOF device was described by Medzihradszky et al., 2000: it is designed for rapid analysis of MALDI-MS and MS/MS spectra, providing high sensitivity, rapid scan speed (10s of spectra per second) and accurate mass capability; truly a high throughput Proteomics analyzer. In addition, the MS/MS provided is high-energy CID, which provides more structurally useful fragmentation than the lower-energy CID MS/MS available on other analyzers. Another TOF-TOF device which involves in-source decay is the so-called LIFT-TOF (Schnaible et al., 2002) which utilizes a different strategy to achieve MS/MS.

Hybrid-MS combinations have been numerous with quadrupole ion traps coupled to TOF. The classical combination of the cylindrical ion trap-TOF (Martin and Brancia, 2003) being developed in the early 1990s and very recently the high-performance linear ion trap has appeared (Hager and Le Blanc, 2003), which is expected to be combined to TOF and FTICR is the near future. The linear ion trap provides an increased volume for ion storage, thus increasing sensitivity and overall performance. As this is a very new device it has only had a limited number of applications primarily directed to smaller mass ions and pharmaceutical applications (Hopfgartner et al., 2003). The potential of FTICR is at last being developed in a true MS/MS combination with both quadrupole ion traps and linear quadrupole ion traps being coupled to FTICR (Thermo Electron, 2003). Hitherto MS/MS in FTICR was usually conducted by sequential mass selection/activation/mass scanning procedures within the single trapping volume. Recent data have shown the truly outstanding sensitivity, high mass capability and accurate mass capability of FTICR, with the main drawbacks being the high price of such systems and future price increases of the liquid helium coolant.

Ion mobility spectrometry is a technique that has largely been restricted to gas/vapour analysis, low cost detection of air pollutants, such as volatile organic compounds, and for explosives screening. It is not a mass detector, but rather detects the shape and size of a molecule, i.e. the cross-section of the species. These devices have been combined with MS and more recently MS/MS analyzers and provide an extra stage of selectivity. Considerable work is currently ongoing with these combinations (Myung et al., 2003; Hoaglund-Hyzer and Clemmer, 2001) and provide selectivity based on the size of the molecule, which can reflect inter alia the degree of folding of a protein. While of the two most common types of IMS analyzer, the classical linear drift analyzer has been used for over 40 years, more recently high-field asymmetric waveform ion mobility spectrometry (FA-IMS) has become available, which is proving popular as an efficient and relatively simple IMS device to couple to mass analyzers (Ells et al., 2000).

It is still uncertain which MS/MS analyzer will prove the most effective when used for high throughput proteomics. Whilst FTICR provides exceptional mass resolution, mass measurement and sensitivity, its performance is very dependent upon the integration time of signal acquisition. It is fundamentally difficult to further improve without very high magnetic field strengths. The greatest improvements will probably occur at the ionization and ion transmission interface into the cyclotron mass analyzer, as this stage is still relatively inefficient with associated ion losses. For proteomics the necessity and utility of such accurate mass measurement is still to be fully determined, although improvement to selectivity should always be advantageous. The future holds great

promise and excitement in the area of instrumentation design, and the application of new spectrometers in proteomic studies.

References

Aebersold, R., 2003. A mass spectrometric journey into protein and proteome research. Journal of the American Society for Mass Spectrometry 14, 685–695.

Andon, N.L., Hollongworth, S., Koller, A., Greenland, A.J., Yates, J.R., Haynes, P.A., 2002. Proteomic characterisation of wheat amyloplasts using identification of proteins by tandem mass spectrometry. Proteomics 2, 1156–1168.

Applied Biosystems, 2002. Achieving reliable protein quantitation results using ICAT reagents, MALDI-TOF mass spectrometry, and advanced data interpretation tool. Application Note 115AP12-01.

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

Bahrman, N., Petit, R.J., 1995. Genetic polymorphism in maritime pine (*Pinus pinaster* Ait.) assessed by two-dimensional gel electrophoresis of needle, bud and pollen proteins. Journal of Molecular Evolution 41, 231–237.

Baldwin, M.A., Medzihradszky, K.F., Lock, C.M., Fisher, B., Settineri, T.A., Burlingame, A.L., 2001. Matrix-assisted laser desorption/ ionization coupled with quadrupole/orthogonal acceleration time-of-flight mass spectrometry for protein discovery, identification, and structural analysis. Analytical Chemistry 73, 1707–1720.

Balmer, Y., Koller, A., del Val, G., Manieri, W., Schurmann, P., Buchanan, B.B., 2003. Proteomics gives insight into the regulatory function of chloroplast thioredoxins. Proceeding of the National Academy of Sciences USA 100, 370–375.

Bardel, J., Louwagie, M., Jaquinod, M., Jourdain, A., Luche, S., Rabilloud, T., Macherel, D., Garin, J., Bourguinon, J., 2002. A survey of the plant mitochondrial proteome in relation to development. Proteomics 2, 880–898.

Bateman, R.H., Carruthers, R., Hoyes, J.B., Jones, C., Langridge, J.I., Millar, A., Vissers, J.P.C., 2002. A novel precursor ion discovery method on a hybrid quadruple orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. Journal of the American Society for Mass Spectrometry 13, 792–803.

Binz, P.A., Hochstrasser, D.F., Appel, R.D., 2003. Mass spectrometry-based proteomics: current status and potential use in clinical chemistry. Clinical Chemistry and Laboratory Medicine 41 (12), 1540–1551.

Black, D.L., 2003. Mechanisms of alternative pre-messenger RNA splicing. Annual Review of Biochemistry 72, 291–336.

Blackburn, R.K., Anderegg, R.J., 1997. Characterization of femtomole levels of proteins in solution using rapid proteolysis and nanoelectrospray ionization mass spectrometry. Journal of the American Society for Mass Spectrometry 8, 483–494.

Bodnar, W.M., Balckburn, R.K., Krise, J.M., Moseley, M.A., 2003. Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. Journal of the American Society for Mass Spectrometry 14, 971–979.

Borederies, G., jamet, E., Lafitte, C., Rossignol, M., Jauneau, A., Boudart, G., Monsarrat, B., Esquerre-Tugaye, M.T., Boudet, A., Pont-Lezica, R., 2003. Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana*: a critical analysis. Electrophoresis 24, 3421–3432.

Borner, G.H.H., Lilley, K.S., Stevens, T.J., Dupree, P., 2003. Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A proteomic and genomic study. Plant Physiology 132, 568–577.

Branlard, G., Amiour, N., Igrejas, A., Gaborit, T., Herbette, S., Dardevet, M., Marion, D., 2003. Diversity of puroindolines as revealed by two-dimensional electrophoresis. Proteomics 3, 168–174.

Bristow, A.W.T., Webb, K.S., 2003. Intercomparison study on accurate mass measurements of small molecules in mass spectrometry. Journal of the American Society for Mass Spectrometry 14 (10), 1086–1098.

Bruins, A.P., 1998. Mechanistic aspects of electrospray ionization. Journal of Chromatography A 794, 345–357.

Butt, Y.K.-C., Lum, J.H.-K., Lo, S.C.-L., 2001. Proteomic identification of plant proteins probed by mammalian nitric oxide synthase antibodies. Planta 216 (5), 762–771.

Bykova, N.V., Egsgaard, H., Moller, I., 2003. Identification of 14 new phosphoproteins involved in important plant mitochondrial processes. FEBS Letters 540, 141–146.

Calikowski, T.T., Meulia, T., Meier, I., 2003. A proteomic study of the *Arabidopsis* nuclear matrix. Journal of Cellular Biochemistry 90, 361–378.

Cambridge HealthTech Institute, 2003. Proteomics Survey Report 2003. Cambridge HealthTech Institute, Newton Upper Falls, MA.

Cao, P., Stults, J.T., 1999. Phosphopeptide analysis by on-line immobilized metal-ion affinity chromatography-capillary electrophoresis-electrospray ionization mass spectrometry. Journal of Chromatography A 853, 225–235.

Carpi, A., Di Maira, G., Vedovato, M., Rossi, V., Naccari, T., Floriduz, M., Terzi, M., Filippini, F., 2002. Comparative proteome bioinformatics: identification of a whole complement of putative protein tyrosine kinases in the model flowering plant *Arabidopsis thaliana*. Proteomics 2, 1494–1503.

Chakraborty, A., Regnier, F., 2002. Global internal standard technology for comparative proteomics. Journal of Chromatography A 949, 173–184.

Chaga, G.S., 2001. Twenty-five years of immobilized metal ion affinity chromatography : past, present and future. Journal of Biochemical and Biophysical Methods 49, 313–334.

Chan, J., Mao, G., Lloyd, C., 2003. Proteomic analysis of plant microtubule-associated proteins. Cell Biology International 27, 181.

Chang, W.P., Huang, L.H., Shen, M., Webster, C., Burlingame, A.L., Roberts, J.K.M., 2000. Pattern of protein synthesis and tolerance to anoxia in root tips of maize seedlings acclimated to a low-oxygen environment and identification of proteins my mass spectrometry. Plant Physiology 122, 295–317.

Chapman, J.R., 1996. Protein and peptide analysis by mass spectrometry. In: Methods in Molecular Biology, vol. 61. Humana Press, Totowa, NJ.

Chapman, J.R., 2000. Mass spectrometry of proteins and peptides. In: Methods in Molecular Biology, vol. 146. Humana Press, Totowa, NJ.

Chapman, K., 2002. The ProteinChip Biomarker System from Ciphergen Biosystems: a novel proteomics platform for rapid biomarker discovery and validation. Biochemical Society Transactions 30 (2), 82–87.

Chivas, S., Ndimba, B.K., Simon, W.J., Robertson, D., Yu, X-L., Knox, P.J., Bolwell, P., Slabas, A.R., 2002. Proteomic analysis of the *Arabidopsis thaliana* cell wall. Electrophoresis 23, 1754–1765.

Chong, B.E., Lubman, D.M., Millar, F.R., Rosenspire, A.J., 1999. Rapid screening of protein profiles of human breast cancer lines using non-porous reversed-phase high performance liquid chromatography separation with matrix assisted laser desorption/ ionization time-of-flight mass spectral analysis. Rapid Communications in Mass Spectrometry 13, 1808–1812.

Clauser, K.R., Baker, P., Burlingame, A.L., 1999. Role of accurate mass measurement (±10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Analytical Chemistry 71, 2871–2882.

Consoli, L., lefevre, A., Zivy, M., de Vienne, D., Damerval, C., 2002. QTL analysis and transcriptome variations for dissecting the genetic architecture of complex traits in maize. Plant Molecular Biology 48, 575–581.

Cooks, R.G., 1978. Collision Spectroscopy. Plenum Press, New York.

Cooper, B., Eckert, D., Andon, N.L., Yates III, J.R., Haynes, P.A., 2003. Investigative proteomics: identification of an unknown plant virus from infected plants using mass spectrometry. Journal of the American Society for Mass Spectrometry 14, 736–741.

Cottingham, K., 2003. Hybrid vigor for mass spectrometers. Analytical Chemistry 75, 315A–319A.

Coutre, J.I., Whitelegge, J.P., Gross, A., Turk, E., Wright, E.M., Kaback, H.R., Faull, K.F., 2000. Proteomics on full-length membrane proteins using mass spectrometry. Biochemistry 39, 4237–4242.

Crofts, A.J., Denecke, J., 1998. Calreticulin and calnexin in plants. Trends in Plant Science 3, 396–399.

Davis, M.T., Lee, T.D., 1997. Variable flow liquid chromatography tandem mass spectrometry and the comprehensive analysis of complex protein digest mixtures. Journal of the American Society for Mass Spectrometry 8, 1059–1069.

Davis, M.T., Beierle, J., Bures, E.T., McGinley, M.D., Mort, J., Robinson, J.H., Spahr, C.S., Yu, W., Luethy, R., Patterson, S.D., 2001. Automated LC-LC-MS-MS platform using binary ion-exchange and gradient reversed-phase chromatography for improved proteomic analyses. Journal of Chromatography B 752, 281–291.

Deutscher, M.P., 1990. Guide to protein purification. In: Methods in Enzymology, vol. 182. Academic Press, New York.

Doroshenko, V.M., Laiko, V.V., Taranenko, N.I., Berkout, V.D., Lee, H.S., 2002. Recent developments in atmospheric pressure MALDI mass spectrometry. International Journal of Mass Spectrometry 221, 39–58.

Eklund, D.M., Edqvist, J., 2003. Localisation of non-specific lipid transfer proteins correlate with programmed cell death responses during endosperm degradation in *Euphorbia lagascae* seedlings. Plant Physiology 132, 1249–1259.

Eksigent Technologies, 2003. NanoLC System for Proteomics Microfluidic HPLC. Available from: <http://www.eksigent.com/>.

Ekström, S., Ericsson, D., Önnerfjord, P., Bengtsson, M., Nisson, J., Marko-Varga, G., Laurell, T., 2001. Signal amplification using "Spot-on-a-chip" technology for the identification of proteins via MALDI-TOF MS. Analytical Chemistry 73, 214–219.

Ells, B., Froese, K., Hrudey, S.E., Purves, R.W., Guevremont, R., Barnett, D.A., 2000. Detection of microcystins using electrospray ionization high-field asymmetric waveform ion mobility mass spectrometry/mass spectrometry. Rapid Communications in Mass Spectrometry 14, 1538–1542.

Eubel, H., Jansch, L., Braun, H.P., 2003. New insights into the respiratory chain of plant mitochondria. Supercomplexes and a unique composition of complex II. Plant Physiology 133, 274–286.

Ferrige, A.G., Seddon, M.J., Skilling, J., Ordsmith, N., 1992. The application of Maxent to high-resolution mass-spectrometry. Rapid Communications In Mass Spectrometry 6, 765–770.

Ferro, M., Salvi, D., Brugiere, S., Miras, S., Kowalski, S., Louwagie, M., Garin, J., Joyard, J., Rolland, N., 2003. Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. Molecular and Cellular Proteomics 2, 325–345.

Ferro, M., Salvi, D., Riviere-Rolland, H., Vermat, T., Seigneurin-Berny, D., Grunwald, D., Garin, J., Joyard, J., Rolland, N., 2002. Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters. Proceedings of the National Academy of Sciences of the United States of America 99, 11487–11492.

Figeys, D., 2003. Proteomics in 2002: a year of technical development and wide ranging applications. Analytical Chemistry 75, 2891–2905.

Froelich, J.E., Wilkerson, C.G., Ray, W.K., McAndrews, R.S., Osteryoung, K.W., Gage, D.A., Phinney, B.S., 2003. Proteomic study of the *Arabidopsis thaliana* chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. Journal of Proteome research 2, 413–425.

Fukao, Y., Hayashi, M., Hara-Nishimura, I., Nishimura, M., 2003. Novel glyoxysomal proein kinase, GPK1, identified by proteomic analysis of glyoxysomes in etiolated cotyledons of *Arabidopsis thaliana*. Plant and Cell Physiology 44, 1002–1012.

Gaberc-Porekar, V., Menart, V., 2001. Perspectives of immobilized-metal affinity chromatography. Journal of Biochemical and Biophysical Methods 49, 335–360.

Gallardo, K., Job, C., Groot, S.P.C., Puype, M., Demol, H., Vandekerckhove, J., Job, D., 2001. Proteomic analysis of *Arabidopsis thaliana* seed germination and priming. Plant Physiology 126, 835–849.

Gallardo, K., Job, C., Groot, S.P.C., Puype, M., Demol, H., Vandekerckhove, J., Job, D., 2002. Proteomics of seed germination. A comparative study of wild-type and gibberellin-deficient seeds. Plant Physiology 129, 823–837.

Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R.D., Burstin, J., 2003. Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiology 133, 664–682.

Gaskell, S.J., 1997. Electrospray: principles and practice. Journal of Mass Spectrometry 32, 677–688.

Giege, P., Sweetlove, L.J., Leaver, C.J., 2003. Identification of mitochondrial protein complexes in *Arabidopsis* using two-dimensional blue-native polyacrylamide gel electrophoresis. Plant Molecular Biology Reports 21, 133–144.

Gluckmann, M., Pfenninger, A., Kruger, R., Thierolf, M., Karas, M., Horneffer, V., Hillenkamp, F., Strupat, K., 2001. Mechanisms in MALDI analysis: surface interaction or incorporation of analytes? International Journal of Mass Spectrometry 210/211, 121–132.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R.L., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchinson, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J.P., Miguel, T., Paszkowski, U., Zhang, S.P., Colbert, M., Sun, W.L., Chen, L.L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y.S., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., Briggs, S., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science 296, 92–100.

Gomez, S., Nishio, J., Faull, K.F., Whitelegge, J.P., 2002. The chloroplast grana proteome defined by intact mass measurements from liquid chromatography mass spectrometry. Molecular and Cellular Proteomics 1, 46–49.

Gomez, S.M., Bil', K.Y., Aguilera, R., Nishio, J.N., Faull, K.F., Whitelegge, J.P., 2003. Transit peptide cleavage sites of integral thylakoid membrane proteins. Molecular and Cellular Proteomics 2, 1068–1085.

Goodlett, D.R., Keller, A., Watts, J.D., Newitt, R., Yi, E.C., Purvine, S., Eng, J.K., Haller, P., Aebersold, R., Kolker, E., 2001. Differential stable isotope labelling of peptides for quantitation and de novo sequence derivation. Rapid Communications in Mass Spectrometry 15, 1214–1221.

Gorg, A., Obermaier, C., Boguth, G., Rarder, A., Schelbe, B., Wildgruber, R., Weiss, W., 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 21, 1037–1053.

Griffin, T.J., Han, D.K.M., Gygi, S.P., Rist, B., Lee, H., Aebersold, R., Parker, K.C., 2001. Toward a high-throughput approach to quantitative proteomic analysis: expression-dependent protein

identification by mass spectrometry. Journal of the American Society for Mass Spectrometry 12, 1238–1246.
Griffiths, W.J., Jonsson, A.P., Liu, S., Rai, D.K., Wang, Y., 2001. Electrospray and tandem mass spectrometry in biochemistry. Journal of Biochemistry 355, 545–561.
Guilhaus, M., 1995. Principles and instrumentation in time-of-flight mass-spectrometry. physical and instrumental concepts. Journal of Mass Spectrometry 30, 1519–1532.
Hager, J.W., Le Blanc, J.C.Y., 2003. High-performance liquid chromatography-tandem mass spectrometry with a new quadrupole/linear ion trap instrument. Journal of Chromatography A 1020, 3–9.
Hale, J.E., Butler, J.P., Knierman, M.D., Becker, G.W., 2000. Increased sensitivity of tryptic peptides detection by MALDI-ToF mass spectrometry is achieved by conversion of lysine to homoarginine. Analytical Biochemistry 287, 110–117.
Hamdam, H., Righetti, P.G., 2002. Modern strategies for protein quantitation in proteome analysis: advantages and limitations. Mass Spectrometry Reviews 21, 287–302.
Hanson, B.J., Schulenberg, B., Patton, W.F., Capaldi, R.A., 2001. A novel subfractionation approach for mitochondrial protein: a three dimensional mitochondrial protein map. Electrophoresis 22, 950–959.
Harris, W.A., Dariusz, J.J., Reilly, J.P., 2002. Use of matrix clusters and trypsin autolysis fragments as mass calibrants in matrix assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 16, 1714–1722.
Haslam, R.P., Downie, A.L., Raveton, M., Gallardo, K., Job, D., Pallett, K.E., John, P., Parry, M.A.J., Coleman, J.O.D., 2003. The assessment of enriched apoplastic extracts using proteomic approaches. Annals of Applied Biology 143, 81–91.
Heazlewood, J.L., Howell, K.A., Whelan, J., Millar, A.H., 2003a. Towards an analysis of the rice mitochondrial proteome. Plant Physiology 132, 230–242.
Heazlewood, J.L., Whelan, J., Millar, A.H., 2003b. The products of the mitochondrial *orf25* and *orfB* genes are F_0 components in the plant F_1F_0 ATP synthase. FEBS Letters 540, 201–205.
Herald, V.L., Heazlewood, J.L., Day, D.A., Millar, A.H., 2003. Proteomic identification of divalent metal cation binding proteins in plant mitochondria. FEBS Letters 537, 96–100.
Hoaglund-Hyzer, C.S., Clemmer, D.E., 2001. Ion trap/ion mobility/quadrupole/time-of-flight mass spectrometry for peptide mixture analysis. Analytical Chemistry 73, 177–184.
Hopfgartner, G., Husser, C., Zell, M., 2003. Rapid screening and characterization of drug metabolites using a new quadrupole-linear ion trap mass spectrometer. Journal of Mass Spectrometry 38, 138–150.
Hoving, S., Vashal, H., van Oostrum, J., 2000. Towards high performance two-dimensional gel electrophoresis using ultrazoom gels. Electrophoresis 21, 2617–2621.
Hoyes, J., Millar, A., Tyldesley, R., Gostick, D., Brown, J., Langridge, J., Ma, A., Choudhary, J., 2002. An investigation into the dynamic range of protein identification; MALDI Vs electrospray on a Q-TOF mass spectrometer. Waters Poster reprint from 5th Siena MS Proteomics Meeting, Siena, Italy.
Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., Hood, L., 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292, 929–934.
Islam, N., Woo, S-H., Tsujimoto, H., Kawasaki, H., Hirano, H., 2002. Proteome approach to characterize seed storage proteins related to ditelocentric chromosomes in common wheat (*Triticum aestivum* L.). Proteomics 2, 1146–1155.
Jansen, R.C., Nap, J-P., Mylnarova, L., 2002. Errors in genomics and proteomics. Nature 20, 19.
Jebanathirajah, J.A., Andersen, S., Blagoev, B., Roepstorff, P., 2002. A rapid screening method to monitor expression of recombinant proteins from various prokaryotic and eukaryotic expression systems using matrix-assisted laser desorption ionization-time-of-flight mass spectrometry. Analytical Biochemistry 305, 242–250.
Joyard, J., Teyssier, E., Miege, C., Berny-Seigneurin, D., Marechal, E., Block, M.A., Dorne, A.J., Rolland, N., Ajlani, G., Douce, R., 1998. The biochemical machinery of plastid envelope membranes. Plant Physiology 118, 715–723.
Kalume, D.E., Molina, H., Pandey, A., 2003. Tackling the phosphoproteome: tools and strategies. Current Opinion in Biotechnology 7, 64–69.
Kannicht, C., 2002. Posttranslational modifications of proteins. In: Methods in Molecular Biology, vol. 194. Humana Press, Totowa, NJ.
Kehr, J., Haebel, S., Blechschmidt-Schneider, S., Willmitzer, L., Steup, M., Fisahn, J., 1999. Analysis of phloem protein patterns from different organs of *Cucurbita maxima* Duch. by matrix-assisted laser desorption/ionization time of flight mass spectroscopy combined with sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Planta 207, 612–619.
Kersten, B., Burkle, L., Kuhn, E.J., Giavalisco, P., Konthur, Z., Lueking, A., Walter, G., Eickhoff, H., Schneider, U., 2002. Large-scale plant proteomics. Plant Molecular Biology 48, 133–141.
Koller, A., Washburn, M.P., Lange, B.M., Andon, N.L., Decieu, C., Haynes, P.A., Hays, L., Schieltz, D., Ulasek, R., Wei, J., Wolters, D., Yates, J.R., 2002. Proteomic survey of metabolic pathways in rice. Proceedings of the National Academy of Sciences of the United States of America 99, 11969–11974.
Komatsu, S., Konishi, H., Shen, S., Yang, G., 2003. Rice proteomics. A step towards functional analysis of the rice genome. Molecular and Cellular Proteomics 2, 2–10.
Koo, A.J., Ohirogge, J.B., 2002. The predicted candidates of *Arabidopsis* plastid inner envelope proteins and their expression profiles. Plant Physiology 130, 823–836.
Kruft, V., Eubel, H., Jansch, L., Werhan, W., Braun, H.P., 2001. Proteomics approach to identify novel proteins in *Arabidopsis*. Plant Physiology 127, 1694–1710.
Kubis, S., Baldwin, A., Patel, R., Razzaq, A., Dupree, P., Lilley, K., Kurth, J., Leister, D., Jarvis, P., 2003. The *Arabidopsis ppi1* mutant is specifically defective in the expression, chloroplast import and accumulation of photosynthetic proteins. The Plant Cell 15, 1859–1871.
Laiko, V.V., Moyer, S.C., Cotter, R.J., 2000. Atmospheric pressure MALDI/ion trap mass spectrometry. Analytical Chemistry 72, 5239–5243.
Larsen, M.R., Cordwell, S.J., Roepstorff, P., 2002. Graphite powder as an alternative or supplement to reversed-phase material for desalting and concentration of peptide mixtures prior to matrix-assisted laser desoption/ionisation mass spectrometry. Proteomics 2, 1277–1287.
Laugesen, S., Roepstorff, P., 2003. Combination of two matrices results in inproved perfromance of MALDI MS for peptide mass mapping and protein analysis. Journal of the American Society for Mass Spectrometry 14, 992–1002.
Lee, H., Yi, E.C., Wen, B., Reily, T.P., Pohl, L., Nelson, S., Aebersold, R., Goodlett, D.R., 2003. Optimization of reversed-phase microcapillary liquid chromatography for quantitative proteomics. Journal of Chromatography B 803 (1), 101–110.
Lee, Y.-S., Mrksich, M., 2002. Protein chips: from concept to practice. Trends in Biotechnology 20 (12), S14–S18.
Lefkovits, I., 2003. Functional and structural proteomics: a critical appraisal. Journal of Chromatography B 787, 1–10.
Leister, D., 2003. Chloroplast research in the genomic age. Trends in Genetics 19, 47–56.
Liebler, D.C., 2002. Introduction to Proteomics: Tools for the New Biology. Humana Press, Totowa, NJ.

Lilley, K.S., Razzaq, A., Dupree, P., 2001. Two-dimensional electrophoresis: recent advances in sample preparation, detection and quantitation. Current Opinion in Chemical Biology 6, 46–50.

Lim, H., Eng, J., Yates III, J.R., Tollaken, S.L., Giometti, C.S., Holden, J.F., Adams, M.W.W., Reich, C.I., Olsen, G.J., Hays, L.G., 2003. Identification of 2D-gel proteins: A comparison of MALDI/TOF peptide mass mapping to μLC-ESI tandem mass spectrometry. Journal of the American Society for Mass Spectrometry 14, 957–970.

Link, A.J., 1995. 2D-proteomic analysis protocols. In: Methods in Molecular Biology, vol. 112. Humana Press, Totowa, NJ.

Link, A.J. (Ed.), 1999. 2-D Proteome Analysis Protocols. Humana Press, Totowa, NJ.

Lipton, M.S., Paša-Toli, L., Anderson, G.A., Anderson, D.J., Auberry, D.L., Battista, J.R., Daly, M.J., Fredrickson, J., Hixson, K.K., Kostandarithes, H., Masselon, C., Meng Markillie, L., Moore, R.J., Romine, M.F., Shen, Y., Stritmatter, E., Toli, N., Udseth, H.R., Venkateswaran, A., Wong, K-K., Zhao, R., Smith, R.D., 2002. Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. Proceedings of the National Academy of Sciences of the United States of America 99, 11049–11054.

Liu, H., Berger, S.J., Chakraborty, A.B., Plumb, R.S., Cohen, S.A., 2002. Multidimensional chromatography coupled to electrospray ionization time-of-flight mass spectrometry as an alternative to two-dimensional gels for the identification and analysis of complex mixtures of intact proteins. Journal of Chromatography B 782, 267–289.

Lubman, D.M., Kachman, M.T., Wang, H., Gong, S., Yan, F., Hamler, R.L., O'Neil, K.A., Zhu, K., Buchanan, N.S., Barder, T.J., 2002. Two-dimensional liquid separations-mass mapping of proteins from human cancer cell lysates. Journal of Chromatography B 782, 183–196.

Mackun, K., Downard, K.M., 2003. Strategy for identifying protein–protein interactions of gel-separated proteins and complexes by mass spectrometry. Analytical Biochemistry 318, 60–70.

Majoul, T., Bancel, E., Triboi, E., Hamida, J.B., Branlard, G., 2003. Proteomic analysis of the effect of heat stress on hexaploid wheat grain: characterization of heat-responsive proteins from total endosperm. Proteomics 3, 175–183.

Maltman, D.J., Simon, W.J., Wheeler, C.H., Dunn, M.J., Wait, R., Slabas, A.R., 2002. Proteomic analysis of the endoplasmic reticulum from developing and germinating seeds of castor (*Ricinus communis*). Electrophoresis 23, 626–639.

Mann, M., Hendrickson, R.C., Pandey, A., 2001. Analysis of proteins and proteomes by mass spectrometry. Annual Review of Biochemistry 70, 437–473.

Mann, M., Jensen, O.N., 2003. Proteomic analysis of posttranslational modifications. Nature Biotechnology 21, 255–261.

Marko-Varga, G., Nilsson, J., Laurell, T., 2003. New directions in miniaturization within the proteomics research area. Electrophoresis 24, 3521–3532.

Martin, R.L., Brancia, F.L., 2003. Analysis of high mass peptides using a novel matrix-assisted laser desorption/ionisation quadrupole ion trap time-of-flight mass spectrometer. Rapid Communications in Mass Spectrometry 17, 1358–1365.

McDonald, W.H., Ohi, R., Miyamoto, D.T., Mitchison, T.J., Yates III, J.R., 2002. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. International Journal of Mass Spectrometry 219, 245–251.

McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E.R., Wilson, R.K., Fulton, R., Kucaba, T.A., Wagner-McPherson, C., Barbazuk, W.B., Gregory, S.G., Humphray, S.J., French, L., Evans, R.S., Bethe, l.G., Whittaker, A., Holden, J.L., McCann, O.T., Dunham, A., Soderlund, C., Scott, C.E., Bentley, D.R., Schuler, G., Chen, H.C., Jang, W.H., Green, E.D., Idol, J.R., Maduro, V.V.B., Montgomery, K.T., Lee, E., Millem, A., Emerling, S., Kucherlapati, R., Gibbs, R., Scherer, S., Gorrell, J.H., Sodergren, E., Clerc-Blankenburg, K., Tabor, P., Naylor, S., Garciam, D., de Jong, P.J., Catanese, J.J., Nowak, N., Osoegawa, K., Qin, S.Z., Rowen, L., Madan, A., Dors, M., Hood, L., Trask, B., Friedman, C., Massa, H., Cheung, V.G., Kirsch, I.R., Reid, T., Yonescu, R., Weissenbach, J., Bruls, T., Heilig, R., Branscomb, E., Olsen, A., Doggett, N., Cheng, J.F., Hawkins, T., Myers, R.M., Shang, J., Ramirez, L., Schmutz, J., Velasquez, O., Dixon, K., Stone, N.E., Cox, D.R., Haussler, D., Kent, W.J., Furey, T., Rogic, S., Kennedy, S., Jones, S., Rosenthal, A., Wen, G.P., Schilhabel, M., Gloeckner, G., Nyakatura, G., Siebert, R., Schlegelberger, B., Korenburg, J., Chen, X.N., Fujiyama, A., Hattori, M., Toyoda, A., Yada, T., Park, H.S., Sakaki, Y., Shimizu, N., Asakawa, S., Kawasaki, K., Sasaki, T., Shintani, A., Shimizu, A., Shibuya, K., Kudoh, J., Minoshima, S., Ramser, J., Seranski, P., Hoff, C., Poustka, A., Reinhardt, R., Lehrach, H., 2001. A physical map of the human genome. Nature 409, 934–941.

Medzihradszky, K.F., Campbell, J.M., Baldwin, M.A., Falick, A.M., Juhasz, P., Vestal, M.L., Burlingame, A.L., 2000. The characteristics of peptide collision-induced dissociation using a high performance MALDI-TOF/TOF tandem mass spectrometer. Analytical Chemistry 72, 552–558.

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

Millar, A.H., Sweetlove, L.J., Giege, P., Leaver, C.J., 2001. Analysis of the *Arabidopsis* mitochondrial proteome. Plant Physiology 127, 1711–1727.

Mo, M., Tse, Y.C., Jiang, L., 2003. Organelle identification and proteomics in plant cells. Trends in Biotechnology 21 (8), 331–332.

Morris, H.R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R.S., Hoyes, J., Bateman, R.H., 1996. High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. Rapid Communications in Mass Spectrometry 10, 889–896.

Myung, S., Lee, Y.J., Moon, M.H., Taraszka, J., Sowell, R., Koeniger, S., Hilderbrand, A.E., Valentine, S.J., Cherbas, L., Cherbas, P., Kaufmann, T.C., Miller, D.F., Mechref, Y., Novotny, M.V., Ewing, M.A., Sporleder, C.R., Clemmer, D.E., 2003. Development of high-sensitivity ion trap ion mobility spectrometry time-of-flight techniques: a high-throughput nano-LC-IMS-TOF separation of peptides arising from a *Drosophila* protein extract. Analytical Chemistry 75, 5137–5145.

Nakayama, G.R., 2001. Proteomics and genomics web alert. Current Opinion in Chemical Biology 6, 9–10.

Nedelkov, D., Nelson, R.W., 2001. Analysis of native proteins from biological fluids by biomolecular interaction analysis mass spectrometry (BIA/MS): exploring the limit of detection, identification of non-specific binding and detection of multi-protein complexes. Biosensors and Bioelectronics 16, 1071–1078.

Nuhse, T.S., Stensballe, A., Jensen, O.N., Peck, S.C., 2003. Large-scale analysis of in vivo phosphorylated membrane proteins by immobilised metal ion affinity chromatography and mass spectrometry. Molecular and Cellular Proteomics 2, 1234–1243.

Oda, Y., Huang, K., Croos, F.R., Cowburn, D., Chait, B.T., 1999. Accurate quantitation of protein expression and site specific phosphorylation. Procedures of the National Academy of Science USA 96, 6591–6596.

Opiteck, G.J., Ramirez, S.M., Jorgenson, J.W., Moseley III, M.A., 1998. Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping. Analytical Biochemistry 258, 349–361.

Patton, W.F., Schulenberg, B., Steinberg, T.H., 2002. Two-dimensional gel electrophoresis; better than a poke in the ICAT? Current Opinion in Biotechnology 13, 321–328.

Peck, S.C., Nuhse, T.S., Hess, D., Iglesias, A., Meins, F., Boller, T., 2001. Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. Plant Cell 13, 1467–1475.

Peltier, J-B., Emanuelsson, O., Kalume, D.E., Ytterberg, J., Friso, G., Rudella, A., Liberles, D.A., Soderberg, L., Roepstorff, P., von Heijne, G., van Wijk, K.J., 2002. Central functions of the lumenal and peripheral thylakoid proteome of *Arabidopsis* determined by experimentation and genome-wide prediction. Plant Cell 14, 211–236.

Peltier, J-B., Friso, G., Kalume, D.E., Roepstorff, P., Nilsson, F., Adamska, I., van Wijk, K.J., 2000. Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. Plant Cell 12, 319–341.

Peltier, J-B., Ytterberg, J., Liberles, D.A., Roepstorff, P., van Wijk, K.J., 2001. Identification of a 35-kDa ClpP protease complex with 10 different ClP isoforms in chloroplasts of *Arabidopsis thaliana*. The Journal of Biological Chemistry 276, 16318–16327.

Person, M.D., Lo, H.-H., Towndrow, K.M., Jia, Z., Monks, T.J., Lau, S.S., 2003. Comparative identification of prostanoid inducible proteins by LC-ESI-MS/MS and MALDI-TOF mass spectrometry. Chemical Research in Toxicology 16, 757–767. Plasmon resonance combined with mass spectrometry. Eur. J. Mass Spectrom 7, 385–391.

Preislar, J., Hu, P., Rejtar, T., Moskovets, E., Karger, B.L., 2002. Capillary array electrophoresis-MALDI mass spectrometry using a vacuum deposition interface. Analytical Chemistry 74, 17–25.

Premstaller, A., Oberacher, H., Walcher, W., Timperio, A.M., Zolla, L., Chervet, J-P., Nukhet, C., van Dorsselaer, A., Huber, C.G., 2001. High performance liquid chromatography-electrospray ionization mass spectrometry using monolithic capillary columns for proteomic studies. Analytical Chemistry 73, 2390–2396.

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

Rabilloud, T., 2002. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics 2, 3–10.

Rakwal, R., Agrawal, G.K., 2003. Rice proteomics: current status and future perspectives. Electrophoresis 24, 3378–3389.

Rakwal, R., Komatsu, S., 2000. Role of jasmonate in the rice (*Oryza sativa* L.) self-defense mechanism using proteome analysis. Electrophoresis 21, 2492–2500.

Rejtar, T., Hu, P., Juhasz, P., Campbell, J.M., Vestal, M.L., Preislar, J., Karger, B.L., 2002. Off-Line coupling of high-resolution capillary electrophoresis to MALDI-TOF and TOF/TOF MS. Journal of Proteome Research 1, 171–179.

Roberts, J.K.M., 2002. Proteomics and a future generation of plant molecular biologists. Plant Molecular Biology 48, 143–154.

Rossignol, M., 2001. Analysis of the plant proteome. Current Opinion in Plant Biotechnology 12, 131–134.

Salekdeh, G.H., Siopongco, J., Wade, L.J., Ghareyazie, B., Bennett, J., 2002. Proteomic analysis of rice leaves during drought stress and recovery. Proteomics 2, 1131–1145.

Santoni, V., Vinh, J., Pflieger, D., Sommerer, N., Maurel, C., 2003. A proteomic study reveals novel insights into the diversity of aquaporin forms expressed in the plasma membrane of plant roots. Biochemical Journal 373, 289–296.

Scalf, M., Westphall, M.S., Smith, L.M., 2000. Charge reduction electrospray mass spectrometry. Analytical Chemistry 72, 52–60.

Schleiff, E., Eichacker, L.A., Eckart, K., Becker, T., Mirus, O., Stahl, T., Soll, J., 2003. Prediction of the plant beta-barrel proteome: a case study of the chloroplast outer envelope. Protein Science 12, 748–759.

Schnaible, V., Wefing, S., Resemann, A., Suckau, D., Bu1cker, A., Wolf-Ku1mmeth, S., Hoffmann, D., 2002. Screening for disulfide bonds in proteins by MALDI in-source decay and LIFT-TOF/TOF-MS. Analytical Chemistry 74, 4980–4988.

Schwartz, H., Mitulovic, G., Van Soest, R., Van Gils, M., VerBerkmoes, N., Hettich, R., 2002. Rapid analysis of bacterial, yeast and plant proteomes using 2D nano LC-MS/MS. Pharmagenomics 2, 42–50.

Schwartz, S.A., Reyzer, M.L., Caprioli, R.M., 2003. Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation. Journal of Mass Spectrometry 38, 699–708.

Scigelova, M., Woffendin, G., Ward, M., Byers, H., Hanger, D., 2002. Detection and identification of phosphorylation sites in proteins using LC/MS/MS with neutral fragment loss mapping. Thermo Finnigan Application Report 304.

Seigneurin-Berny, D., Rolland, N., Garin, J., Joyard, J., 1999. Differential extraction of hydrophobic proteins from chloroplast envelope membranes: a subcellular-specific proteomic approach to identify rare intrinsic membrane proteins. Plant Journal 19, 217–228.

Sharma, J., Panico, M., Barber, J., Morris, H.R., 1997. Purification and determination of intact molecular mass by electrospray ionization mass spectrometry of the photosystem II reaction centre subunits. Journal of Biological Chemistry 272, 33153–33157.

Sheil, M.M., Bordoli, R.S., Langridge, J., 2001. Femtomole sequence analysis of a mixed protein digest with automated data interpretation using a Q-TOF mass spectrometer. Application note 229 from Micromass.

Shen, S., Jing, Y., Kuang, T., 2003a. Proteomics approach to identify wound-response related proteins from rice leaf sheath. Proteomics 3, 527–535.

Shen, S., Sharma, A., Komatsu, S., 2003b. Characterisation of proteins responsive to gibberellin in the leaf-sheath of rice (*Oryza sativa* L.) seedling using proteome analysis. Biological Pharmacology Bulletin 26, 129–136.

Shen, Y., Tolic, N., Masselson, C., Pasa-Tolic, L., Camp II, D.G., Hixson, K.K., Zhao, R., Anderson, G.A., Smith, R.D., 2004. Ultrasensitive proteomics using high-efficiency on-line Micro-SPE-Nano-LC-NanoESI MS and MS/MS. Analytical Chemistry 76, 144–154.

Shevchenko, A., Matthias, W., Vorm, O., Mann, M., 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Analytical Chemistry 68, 850–858.

Skylas, D.J., Copeland, L., Rathmell, W.G., Wrigley, C.W., 2001. The wheat grain as a basis for more efficient cultivar identification. Proteomics 1, 1542–1546.

Sobott, F., Hernandez, H., McCammon, M.G., Tito, M.A., Robinson, C.R., 2002. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. Analytical Chemistry 74, 1402–1407.

Sonksen, C.P., Roepstorff, P., Markgren, P.O., Danielson, U.H., Hamalainen, M.D., Jansson, O., 2001. Capture and analysis of low molecular weight ligands by surface.

Stancato, L.F., Petricoin, E.F., 2001. Fingerprinting of signal transduction pathways using a combination of antiphosphotyrosine immunoprecipitations and two-dimensional polyacrylamide gel electrophoresis. Electrophoresis 22, 2120–2124.

Stevanovic, S., Bohley, P., 2001. Proteome analysis by three dimensional protein separation: turnover of cytosolic proteins in hepatocytes. Biological Chemistry 382, 677–682.

Stroebel, D., Choquet, Y., Popot, J-L., Picot, D., 2003. An atypical heam in the cytochrome b_6f complex. Nature 426, 413–418.

Sweetlove, L.J., Heazlewood, J.L., Herald, V., Holtzapffel, R., Day, D.A., Leaver, C.J., Millar, A.H., 2002. The impact of oxidative stress on *Arabidopsis* mitochondria. Plant Journal 32, 891–904.

Tang, K., Shahgholi, M., Garcia, B.A., Heaney, P.J., Cantor, C.R., 2002. Improvement in the apparent mass resolution of oligonucleotides by using $^{12}C/^{14}N$-enriched samples. Analytical Chemistry 74, 226–231.

Taylor, S.W., Fahly, E., Ghosh, S.S., 2003. Global organellar proteomics. Trends in Biotechnology 21, 82–88.

ThermoElectron Corp., 2003. Available from: <http://www.thermo-bremen.com/products/ftms/ltq-ft_main.htm>.

Thiellement, H., Zivy, M., Plomion, C., 2002. Combining proteomic and genetic studies in plants. Journal of Chromatography B 782, 137–149.

Thiellement, H., Bahrman, N., Damerval, C., Plomion, C., Rossignol, M., Santoni, V., de Vienne, D., Zivy, M., 1999. Proteomics for genetic and physiological studies in plants. Electrophoresis 20, 2013–2026.

Valaskovic, G.A., Kelleher, N.L., 2002. Miniaturized formats for efficient mass spectrometry-based proteomics and therapeutic development. Current Topics in Medicinal Chemistry 2 (1), 1–12.

van Wijk, K.J., 2000. Proteomics of the chloroplast: experimentation and prediction. Trends in Plant Science 5, 421–425.

Vener, A.V., Harms, A., Sussman, M.R., Vierstra, R.D., 2001. Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of *Arabidopsis thaliana*. The Journal of Biological Chemistry 276, 6959–6966.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., Bork, P., 2002. Comparative assessment of large-scale data sets of protein–protein interactions. Nature 417, 399–403.

Wagner, Y., Sickmann, A., Meyer, H.E., Daum, G., 2003. Multidimensional nano-HPLC for the analysis of protein complexes. Journal of the American Society for Mass Spectrometry 14, 1003–1011.

Walcher, W., Oberacher, H., Troiani, S., Holzl, G., Oefner, P., Zolla, L., Huber, C.G., 2002. Monolithic capillary columns for liquid chromatography-electrospray ionization mass spectrometry in proteomic and genomic research. Journal of Chromatography B 782, 111–125.

Walker, J.M., 2002. The Protein Protocols Handbook. Humana Press, Totowa, NJ.

Wall, D., Kachman, M.T., Gong, S., Hinderer, R., Parus, S., Misek, D., Hanash, S., Lubman, D.M., 2000. Isoelectric focusing nonporous RP HPLC: a two-dimensional liquid-phase separation method for mapping of cellular proteins with identification using MALDI-TOF mass spectrometry. Analytical Chemistry 72, 1099–1111.

Wang, S., Regnier, F.E., 2001. Proteolysis of whole cell extracts with immobilized enzyme columns as part of multidimensional chromatography. Journal of Chromatography A 913, 429–436.

Wang, S., Zhang, X., Regnier, F., 2002. Quantitaive proteomics strategy involving the selection of peptides containing both cytseine and histidine from tryptic digests of cell lysates. Journal of Chromatography A 949, 153–162.

Wasinger, V.C., Cordwell, X.J., Cerpapoljak, A., Yan, O.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, K.L., Humphrey-Smith, I., 1995. Progress with gene-product mapping of the molliculites – *Mycoplasm genitalium*. Electrophoresis 16, 1090–1094.

Watson, B.S., Asrivatham, V.S., Wang, L., Sumner, L.W., 2003. Mapping the proteome of Barrel Medic. Plant Physiology 131, 1104–1123.

Westermeier, R., Naven, T., 2002. Proteomics in Practice. Wiley, Weinheim, Germany.

Whitelegge, J.P., 2003. Thylakoid membrane proteomics. Photosynthesis Research 78, 265–277.

Whitelegge, J.P., Gunderson, C.B., Faull, K.F., 1998. Electrospray-inoization mass spectrometry of intact intrinsic proteins. Protein Science 7, 1423–1430.

Whitelegge, J.P., Zhang, H., Aguilera, R., Taylor, R.M., Cramer, W.A., 2002. Full subunit coverage liquid chromatography electrospray mass spectrometry (LCMS+) of an oligomeric membrane protein. Molecular and Cellular Proteomics 1, 816–827.

Williams, J.D., Flanagan, M., Lopez, L., Fischer, S., Miller, L.A.D., 2003. Using accurate mass electrospray ionization time-of-flight mass spectrometry with in-source collision-induce dissociation to sequence peptide mixtures. Journal of Chromatography A 1020, 11–26.

Wilson, K.A., McManus, M.T., Gordon, M.E., Jordan, T.W., 2002. The proteomics of senescence in leaves of white clover *Trifolium repens* (L.). Proteomics 2, 1114–1122.

Witters, E., Laukens, K., Deckers, P., Van Dongen, W., Esmans, E., Van Onckelen, H., 2003. Fast liquid chromatography coupled to electrospray tandem mass spectrometry peptide sequencing for cross-species protein identification. Rapid Communications in Mass Spectrometry 17, 2188–2194.

Wolff, J.-C., Eckers, C., Sage, A.B., Giles, K., Bateman, R., 2001. Accurate mass liquid chromatography/mass spectrometry on quadrupole orthogonal acceleration time-of-flight mass analyzers using switching between separate sample and reference sprays. 2. Applications using the dual-electrospray ion source. Analytical Chemistry 73, 2605–2612.

Wolters, D.A., Washburn, M.P., Yates 3rd., J.R., 2001. An automated multidimensional protein identification technology for shotgun proteomics. Analytical Chemistry 73, 5683–5690.

Wong, J.H., Balmer, Y., Cai, N., Tanaka, C.K., Vensel, W.H., Hurkmam, W.J., Buchanan, B.B., 2003. Unraveling thioredoxin-linked metabolic processes of cereal starchy endosperm using proteomics. FEBS Letters 547, 151–156.

Wu, C.W., Yates, J.R., 2003. The application of mass spectrometry to membrane proteomics. Nature Biotechnology, 21262–21267.

Xing, T., Rampitsch, C., Muki, B.L., Mauthe, W., Stebbing, J.A., Malik, K., Jordan, M., 2003. MALDI-Q-TOF-MS and transient expression analysis indicated co-enhancement of ⋉-1,3-glucanase and endochitinase by tMEK2 and the involvement of divergent pathways. Physiological and Molecular Plant Pathology 62, 209–217.

Xu, B.G.J., Caprioli, R.M., 2002. Direct analysis of laser capture microdissected cells by MALDI mass spectrometry. Journal of the American Society for Mass Spectrometry 13, 1292–1297.

Yamaguchi, K., Subramanian, A.R., 2000. The plastid ribosome; identification of all the proteins in the 50 S subunit of an organelle ribosome (chloroplast). Journal of Biological Chemistry 275, 28466–28482.

Yao, X., Freas, A., Ramirez, J., Demirev, P.A., Fenselau, C., 2001. Proteolytic 18O labelling for comparative proteomics: model studies with two serotypes of adenovirus. Analytical Chemistry 73, 2836–2842.

Yu, J., Hu, S.N., Wang, J., Wong, G.K.S., Li, S.G., Liu, B., Deng, Y.J., Dai, L., Zhou, Y., Zhang, X.Q., Cao, M.L., Liu, J., Sun, J.D., Tang, J.B., Chen, Y.J., Huang, X.B., Lin, W., Ye, C., Tong, W., Cong, L.J., Geng, J.N., Han, Y.J., Li, L., Li, W., Hu, G.Q., Huang, X.G., Li, W.J., Li, J., Liu, Z.W., Li, L., Liu, J.P., Qi, Q.H., Liu, J.S., Li, L., Li, T., Wang, X.G., Lu, H., Wu, T.T., Zhu, M., Ni, P.X., Han, H., Dong, W., Ren, X.Y., Feng, X.L., Cui, P., Li, X.R., Wang, H., Xu, X., Zhai, W.X., Xu, Z., Zhang, J.S., He, S.J., Zhang, J.G., Xu, J.C., Zhang, K.L., Zheng, X.W., Dong, J.H., Zeng, W.Y., Tao, L., Ye, J., Tan, J., Ren, X.D., Chen, X.W., He, J., Liu, D.F., Tian, W., Tian, C.G., Xia, H.G., Bao, Q.Y., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W.M., Li, P., Chen, W., Wang, X.D., Zhang, Y., Hu, J.F., Wang, J., Liu, S., Yang, J., Zhang, G.Y., Xiong, Y.Q., Li, Z.J., Mao, L., Zhou, C.S., Zhu, Z., Chen, R.S., Hao, B.L., Zheng, W.M., Chen, S.Y., Guo, W., Li, G.J., Liu, S.Q., Tao, M., Wang, J., Zhu, L.H., Yuan, L.P., Yang, H.M., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). Science 296, 79–92.

Zabrouskov, V., Giacomelli, L., van Wijk, K.J., McLafferty, F.W., 2003. A new approach for plant proteomics: characterization of

chloroplast proteins of *Arabidopsis thaliana* by top-down mass spectrometry. Molecular and Cellular Proteomics 2, 1253–1260.

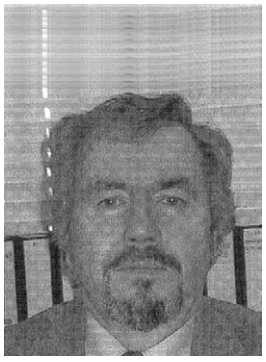
Zhang, Z., McElvain, J.S., 2000. De novo peptide sequencing by two-dimensional fragment correlation mass spectrometry. Analytical Chemistry 72, 2337–2350.

Zheleva, D., Sharma, J., Panico, M., Morris, H.R., Barber, J., 1998. Isolation and characterization of monomeric and dimeric CP47-reaction center photosystem II complexes. Journal of Biological Chemistry 273, 16122–16127.

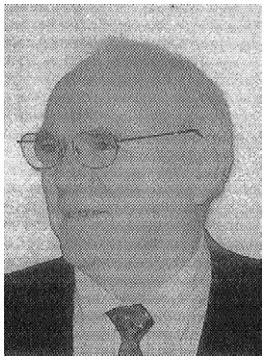
Zhu, H., Bilgin, M., Snyder, M., 2003. Proteomics. Annual Review of Biochemistry 72, 783–812.

Zivy, M., de Vienne, D., 2000. Proteomics: a link between genomics, genetics and physiology. Plant Molecular Biology 44, 575–580.

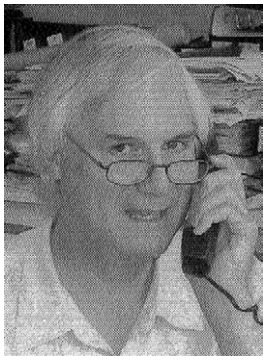
Zolla, L., Rinalducci, S., Timperio, A.M., Huber, C.G., 2002. Proteomics of light-harvesting proteins in different plant species. Analysis and comparison by liquid chromatography–electrospray ionization mass spectrometry. Photosystem I. Plant Physiology 130, 1938–1950.



Russ Newton is Professor of Biochemistry at the University of Wales Swansea. A native of Merseyside, he obtained his BSc in Biochemistry at the University of Liverpool in 1968, and also his PhD in 1971 for studies of Vitamin E Biochemistry in plants. He joined the University College Swansea as a postdoctoral researcher in 1971 and progressed through the ranks to be awarded a Personal Chair in 2000. In 1996 he was awarded a DSc by the University of Liverpool for ''an outstanding contribution to knowledge of cyclic nucleotide biochemistry''; he also has Fellowships of the Royal Society of Chemistry and of the Institute of Analysts and Programmers. In 2000 he was a Visiting Professor at the University of Antwerp, where he retains active and productive collaborations with Professors Eddy Esmans (Mass Spectrometry) and Harry van Onckelen (Plant Biochemistry). In addition to being Head of the Biochemistry Group at UWS he is Lead Director of the Biomolecular Analysis Mass Spectrometry Facility: major research interests are the biochemistry of second messengers in both plants and animals, and the development and application of modern mass spectrometric methods in biomolecular analysis.



Gareth Brenton is a native of west Wales; his mass spectrometry career started as a postdoctoral fellow in Professor John Beynon's Royal Society Research Unit in the University of Wales Swansea. Industrial experience was obtained with AEI-Kratos (Manchester, UK), first as a scientific engineer then as a product leader of their high-resolution mass spectrometry range. He took up a mass spectrometry position in Swansea in 1982, where he researched into the fundamentals of ion/molecule collision spectroscopy. His research interests are in the area of instrument design and development, analytical chemistry and biomolecular applications. Since 1984 he has collaborated with the Biochemistry Research Group and was promoted to a Personal Chair in Mass Spectrometry in 1996. He is the first recipient of the Curt Brunnée Research Award 'For significant contributions to the development of instrumentation in mass spectrometry' which was presented in Budapest at the 13th International MS Conference. He was an Editor of Rapid Communications in Mass Spectrometry for over 15 years and is on the editorial board of other mass spectrometry journals. He is the current Chairman of the British Mass Spectrometry Society and is a co-founder of Microptic Ltd which invented and markets an automated Cytogenetic screening system used to identify hazardous chemicals/pharmaceuticals. He is the Director of the EPRSC National Mass Spectrometry Service Centre, located in Swansea, which runs around 20,000 samples for the UK university chemistry research community.



Chris Smith, Originally from Birmingham, is a graduate of the University of Wales Aberystwyth where he was awarded his BSe in 1970. He subsequently carried out research into the structure and metabolism of plant cell walls in Professor Mike Hall's lab in Aberystwyth, and was awarded his PhD from the University of Wales in 1974. In the same year he moved to Professor Eric Brown's lab at the Department of Biochemistry at University of Wales Swansea to work on plant cyclic nucleotide biochemistry. It was there that his collaboration with Professor Russ Newton began. He is currently a senior lecturer in the School of Biological Sciences at Swansea and an associate member of the Biomolecular Analysis Mass Spectrometry Laboratory. For several years Chris was a member of the committee of the Phytochemical Society of Europe and served as honorary secretary of the Committee. He has also served on the Wales & West Professional & Educational committee of the Biochemical Society, actively promoting scientific meetings for postgraduate members. His current research interests are in the area of intracellular signalling in plants and the cyanobacterium *Gloeothece*.

Dr Ed Dudley is the Experimental Officer of the Biomolecular Analysis Mass Spectrometry (BAMS) Laboratory, a collaborative facility serving the School of Biological Sciences, the Department of Chemistry and the Clinical School of the University of Wales Swansea. His PhD involved the application of HPLC-MS to the identification and evaluation of the urinary modified nucleosides as potential cancer markers. After two years working, again with HPLC-MS, in the drug analysis development and validation industry he took up his present position. His current collaborative projects include the study of biologically important cyclic peptides and their degradation products, fragmentations of nucleosides, nucleotides and cyclic nucleotides and their analogues, the study of the role of ubiquitinone in protein-mediated drug transportation, the purification and identification of a novel antimicrobial factor with anti-MRSA activity, qualitative and quantitative analysis of cyclic nucleotide-initiated phosphorylations, the structural study of the biopolymer Sporopollenin and the development of a miniaturised HPLC-MS method for assay of guanine deaminase activity as an indicator of liver function.