CMLS, Cell. Mol. Life Sci. 62 (2005) 848–869
1420-682X/05/080848-22
DOI 10.1007/s00018-005-5006-6

CMLS Cellular and Molecular Life Sciences

Review

Mass spectrometry-based proteomics in the life sciences

C. S. Lane

Department of Pharmaceutical and Biological Chemistry, The School of Pharmacy, University of London, 29/39 Brunswick Square, London WC1N 1AX (United Kingdom), Fax: + 44 20 7753 5964, e-mail: catherine.lane@ulsop.ac.uk

Received 7 January 2005; accepted 27 January 2005

Abstract. Over the last 20 years, mass spectrometry-based proteomics has become an indispensable tool in the cellular and molecular life sciences. This has been enabled by the 'soft ionisation' techniques of electrospray and matrix-assisted laser desorption-ionisation, which allow the gentle ionisation and vaporisation of large, thermally labile biomolecules. Innovative instrumentation designs and biochemical strategies have brought success in the large-scale identification and quantification of proteins, as well as the characterisation of their complexes and post-translational modifications. This review describes the instrumentation used for proteomics research. It presents an overview of the current applications of mass spectrometry-based proteomics to the cellular and molecular life sciences, and discusses challenges that exist for research in the future.

Key words. Mass spectrometry; proteomics; proteins; peptides; electrospray; MALDI; post-translational modifications; protein interactions.

Introduction

The genomic sequencing of numerous organisms has transformed biological and medical research, providing the foundation for the large-scale interpretation of gene and cellular function. The term proteome was coined in 1994 to describe the set of proteins encoded by the genome [1]. Proteomics, the study of the proteome, has come to encompass the identification, characterisation and quantification of the complete set of proteins expressed by the entire genome in the lifetime of a given cell, tissue or organism, including isoforms, polymorphisms and modifications, protein-protein interactions and the structural description of proteins and their complexes. Most biological functions are carried out by proteins, and to understand how cells work, one must study the proteins present, what they do and how they interact with one another. If the genome represents the words in a dictionary, then the proteome provides the definitions, with the interactions of the proteins with each other and the other molecules in their environment providing the grammar to form meaningful language.

Although genomics provided the 'blueprint' for the potential gene products that are the focus of proteomic studies, the challenges of proteomics are larger and far more complex than the huge but basically straightforward task of mapping the genome. In contrast to the static nature of the genome, which is essentially identical in every cell of an organism, the proteome is dynamic, constantly changing and responding to internal and external stimuli. Whereas DNA sequencing has the enabling technologies of the polymerase chain reaction and automated sequencing, proteomics must cope with problems of limited and variable sample material, a protein abundance dynamic range of more than 10^6-fold, post-translational modifications and a plethora of perturbations due to development, environment, drugs and disease [2]. The total number of human genes is not known, but has been estimated at around 25,000 [3]; this number of genes could give rise to over two million protein components [4].

The sheer scale and complexity of proteomics research makes it very much a technology-driven enterprise, and one that is rapidly evolving with the invention and development of methods and techniques. A recent *Nature* overview [2] cites five 'central pillars' of proteomics research as being mass spectrometry-based proteomics, array-based proteomics (for example the yeast two-hybrid system), structural proteomics (encompassing the use of X-ray crystallography, nuclear magnetic resonance spectroscopy, electron microscopy and electron tomography), clinical proteomics (i.e. disease states and drug discovery) and proteome informatics. Yet proteomics is difficult to section into different areas: the boundaries between the subjects are blurred, and much proteomics research involves the combination of technologies from several different fields.

Mass spectrometry-based proteomics

Mass spectrometry (MS) is a technique whose beginnings date back to the studies performed by J. J. Thomson [5] and his student F. W. Aston [6] in the early days of the last century. MS measures, with extremely high sensitivity, the mass to charge ratios (*m/z*) of gas-phase ions. In the last 20 years, MS has played an increasingly significant role in the biological sciences, and today is the most sensitive method for the structural characterisation of biomolecules [7]. The successes of MS in biology are largely due to the introduction of the 'soft ionisation' techniques of electrospray (ES) [8–11] and matrix-assisted laser desorption ionisation (MALDI) [12, 13], which allow the transfer of large, polar, thermally labile biomolecules into the gaseous phase for mass analysis, without prior derivatisation. These achievements were recognised by the awarding of the 2002 Nobel prize in chemistry to John Fenn and Koichi Tanaka [14] for their pioneering work on ES and MALDI.

At the same time as ES and MALDI were becoming commonplace in biological MS, the concept of proteomics was emerging. The origins of what has come to be called proteomics date back to the invention of two-dimensional gels 30 years ago, which provided the first practical method for displaying thousands of proteins on a single gel [15, 16]. Biological MS allowed the large-scale identification of those proteins, and has now developed to such an extent as to supersede the two-dimensional gels that originally gave proteomics its impetus.

Instrumentation

MS measures the *m/z* ratios of gaseous ions. Every mass spectrometer now consists of: an ion source, to produce ions from the sample; one or more mass analysers, to separate the ions according to their *m/z* ratios; a detector, to register the number of ions emerging from the last analyser; and a computer, to process the data, to produce the mass spectrum in a suitable form and to control the instrument through feedback. Each mass spectrometer also has an inlet device to introduce the analyte into the ion source, for example a liquid chromatograph or a direct insertion probe (fig. 1).

Ion sources

To analyse a sample by MS, it must first be vaporised and ionised. The two ionisation techniques most commonly used for the mass spectrometric analysis of proteins and peptides are ES and MALDI. ES produces gaseous ions from solution phase samples, and can therefore be easily coupled to liquid-based separation technologies such as liquid chromatography (LC) and capillary electrophoresis. MALDI ionises samples out of a dry, crystalline matrix, and is generally used to analyse simple peptide mixtures, whilst integrated ES-LC-MS systems are favoured for the analysis of more complex samples.

The electrospray process

The generation of macroions by ES was first demonstrated by Dole et al. in 1968 [17], but it was Fenn's group at Yale University that first coupled ES with MS [8, 9]. The ES process (fig. 2) transfers ions in solution into gaseous ions at atmospheric pressure, which are sampled into the vacuum system of the mass spectrometer through a series of sampling apertures separating successive vacuum stages.

The mechanisms involved in the production of isolated gaseous ions by ES are not fully understood [18]. The sample solution flows at low flow rates (nl min^{-1} to µl min^{-1}) through a capillary tube to which a high voltage (1–6 kV) is applied. The solution flowing through the

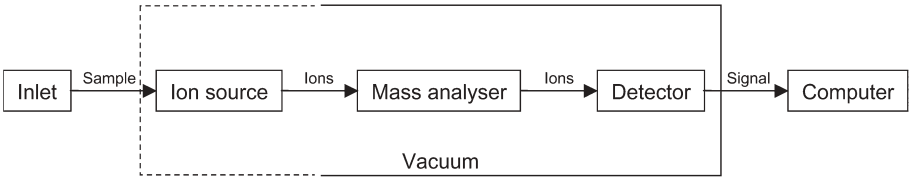

Figure 1. Basic diagram for a mass spectrometer.

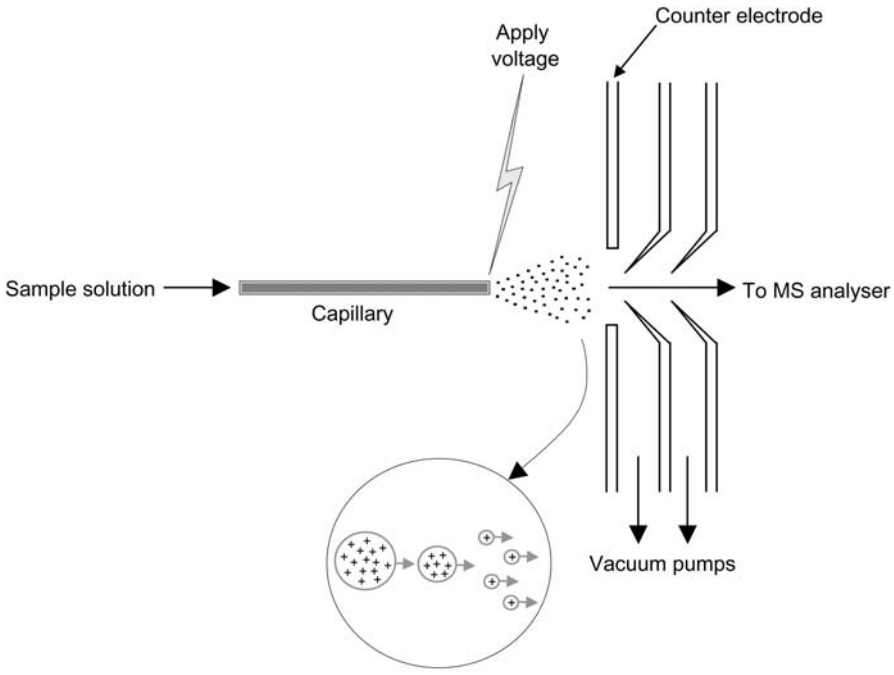

Figure 2. Schematic diagram of an ES interface and the ES process.

capillary experiences an electric field set up between the capillary and a counter electrode and, assuming a positive potential is applied to the capillary, positive ions in solution will accumulate at the surface of the tip, which becomes drawn out, assuming a conical shape known as a 'Taylor cone' (fig. 3). As the liquid is forced to hold more electric charge, the cone is drawn out into a filament that, when the surface tension is exceeded by the applied electrostatic force, produces positively charged droplets (< 10 µm in diameter) via a 'budding' process. Figure 3 shows a photograph of the spray produced. The droplets fly towards the counter electrode (or collector), which is opposite in charge to their own. As they fly towards the electrode they pass through either a heated capillary (180–270°C) or a curtain of heated nitrogen to allow solvent to evaporate. The electrical charge density at the surface of the droplets increases as the droplet size decreases. The droplet deforms into a tear shape and, at the point where the electrostatic repulsion is greater than the surface tension (known as the 'Rayleigh limit'), it blows apart, emitting smaller particles (fig. 4). Depending on the initial size of the droplet, the particles leaving can either be smaller droplets that repeat the process, or discrete solvated surface ions. At atmospheric pressure, collisions with the surrounding gases quickly desolvate the solvent-clustered ion, resulting in a quasi-molecular (or multi-charged) ion [19].

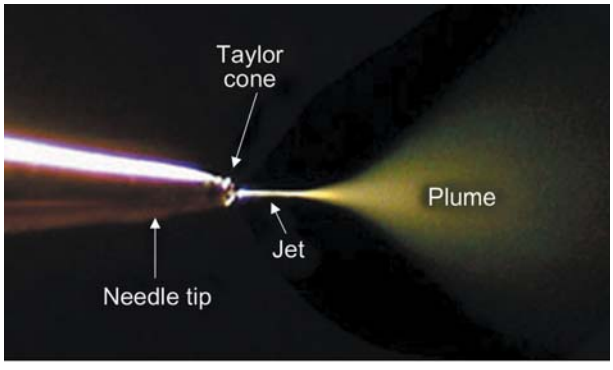

Figure 3. The ES process, viewed through a high-powered microscope. As the liquid begins to exit the needle it charges up and assumes a conical shape, known as the Taylor cone. At the tip of the cone, the liquid is drawn to a filament, which then becomes unstable, breaking up into a mist of charged droplets. Since the droplets are charged they repel each other strongly and fly apart, covering a wide surface area. Taken from http://www.newobjective.com/electrospray/.

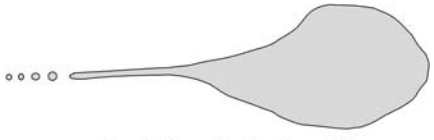

Figure 4. Drawing of a decomposing droplet in an ES source [23]. The critical point at which the surface tension of the droplet is overcome by the electrostatic repulsion of the surface charges is known as the Rayleigh stability limit: q = charge; ε_0 = permittivity of the environment; γ = surface tension and D = diameter of a supposed spherical droplet.

An alternative mechanism for gas phase ion production has been proposed in which ion emission is envisaged to occur from small, highly charged droplets, with the driving force for ion formation being the repulsion between the ion and the other charges on the droplet [20, 21]. The relative importance of the two proposed mechanisms remains the subject of ongoing research and debate [22].

Significantly, the ES process occurs at relatively low temperatures (room temperature, or just above), and so large, thermally labile, polar molecules can be ionised without decomposition. Ionised molecules of the form $[M + H]^+$ or $[M + nH]^{n+}$ (or $[M - H]^-$/$[M - nH]^{n-}$) are generally produced. The prerequisite for gaseous ion production with ES is that the analyte can be ionised in solution. If several ionisable sites are present, multiply charged ions will be produced; for example, denatured proteins typically carry one charge per 1000 Da [23]. By observing such multiply charged species, the effective mass range of the spectrometer can be extended to hundreds of thousands of daltons.

Matrix-assisted laser desorption ionisation

The MALDI technique was introduced in 1988 when Karas and Hillenkamp described the analysis of proteins with molecular masses exceeding 10 kDa [12, 13]. MALDI sublimates and ionises the analyte out of a dry, crystalline matrix using laser pulses.

The sample to be analysed is co-crystallised with a large excess of a matrix material that will strongly absorb the light from a laser. The laser is typically a nitrogen laser at 337 nm. Irradiation of the matrix causes rapid heating and localised sublimation of the matrix crystals. Since the matrix is in large excess and contains a chromophore for the laser light it will absorb essentially all of the laser radiation. As the matrix expands into the gas phase it takes with it intact analyte molecules; little internal energy is transferred to the analyte molecules, allowing ionisation without fragmentation. Ionisation can occur at any time during this process, but the exact origin of ions produced by the MALDI process is still not fully understood. The most widely accepted mechanism involves gas phase proton transfer in the expanding matrix plume with photoionised matrix molecules [23]. This is illustrated in figure 5.

The MALDI process is independent of the absorption properties and size of the compound to be analysed and therefore allows the desorption and ionisation of analytes with very high molecular masses (greater than 100,000 Da).

Mass analysers

To a large extent, the information obtained from a proteomics experiment is determined by the operating conditions and performance specifications of the mass analyser. The mass analyser is the means by which the ions are separated and detected; key instrumental parameters for assessing performance and utility include mass resolution, mass accuracy, mass range, sensitivity and capability for tandem MS (MS/MS or MS^n where n = 2, 3, 4, …; this is the generation of fragment ion spectra from selected precursor ions). There are four basic types of mass analyser currently in use for proteomics research: the ion trap, time-of-flight (TOF), quadrupole and Fourier transform (FT) ion cyclotron resonance (ICR) analysers. They are diverse in terms of design and performance, and can either be used as stand-alone analysers or,

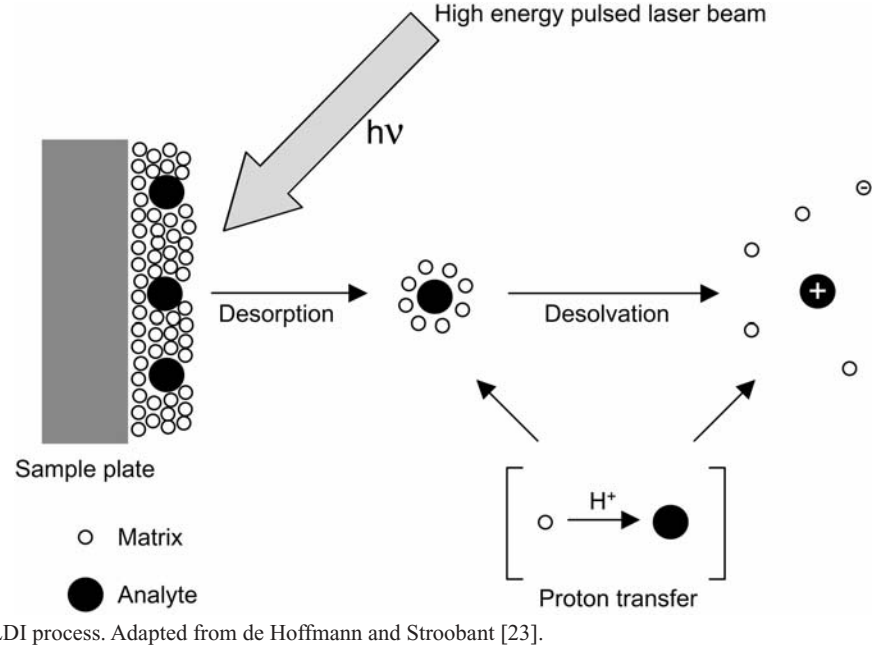

Figure 5. The MALDI process. Adapted from de Hoffmann and Stroobant [23].

in some cases, put together in tandem to take advantage of their different strengths.

Quadrupoles

Quadrupole mass analysers consist of four precisely parallel rods equally spaced around a central axis. Opposing sets of rods have both a dc (direct current) and an ac (alternating current) or rf (radio frequency) voltage component, one set positive and the other set negative. Ions are introduced in a continuous beam along the central axis between the poles (fig. 6), and are filtered on the basis of their *m/z* ratios in the following manner: ions that pass between the two positive rods that are above a critical *m/z* ratio are transmitted through the centre of the quadrupole. This forms a high-pass mass filter. Ions that pass between the two rods with a negative potential that are below a critical *m/z* ratio are transmitted through the centre of the quadrupole; this forms a low pass mass filter. Combining both sets of rods into a quadrupole arrangement overlaps the two mass filter regions, creating a 'band pass' area of mutual stability (fig. 7) and allowing ions of a certain *m/z* ratio to pass through. Ions with *m/z* ratios outside this area of mutual stability cannot pass through and run into the rods. The *m/z* ratio of the ions that are allowed to pass through the quadrupole is proportional to the voltage applied to the rods; the higher the voltage, the higher the *m/z* value that is allowed to pass. By altering the relative contributions of the dc and rf components, the width of the band pass area, and therefore the resolution, can be adjusted (wider band pass = wider peak = lower resolution; narrower band pass = narrower peak = higher resolution). Scanning a quadrupole mass analyser involves ramping the amplitude of the dc and rf voltages at a constant ratio, thus changing the position of the band pass region and allowing different masses to be transmitted.

For MS/MS analysis, three quadrupoles can be configured together (to form a 'triplequad'). The first and third quadrupoles are used for scanning, whilst the middle quadrupole is used as a collision cell. Ions in the second quadrupole are fragmented by collision-activated dissociation (CAD): low-energy collisions with a background gas such as nitrogen.

Ion traps

Ion traps work by first trapping the ions and then detecting them based on their *m/z* ratios. Conceptually, an ion trap can be imagined as a quadrupole bent around on itself to form a closed loop. The inner rod is reduced to a point at the centre of the trap; the outer rod is a circular ring electrode, and the top and bottom rods become two end cap electrodes (fig. 8). Hence the ion trap is also referred to as the 'quadrupole ion trap'. Ions are held inside the trap and subjected to oscillating electric fields generated by an rf voltage applied to the ring electrode only. Unlike quadrupoles, there is no dc component to the voltage, and the end caps are held near ground potential. For mass analysis, ions are guided up to the trap and a portion of the ions is sampled into the trap. Ions of all *m/z* values enter the trap at the same time. Once inside the trap, the ions are confined by the rf field and take on an oscillating frequency that is related to their *m/z* value (fig. 9). As the ions repel each other in the trap their trajectories expand as a function of time. To avoid ion losses from this expansion, helium gas is present inside the trap as a 'buffer' gas, to remove excess energy from the ions by collision; the helium 'dampens' the ions into the centre of the ring electrode.

To scan the mass range, the rf voltage (i.e. amplitude) on the ring electrode is ramped. At the same time, a small rf voltage is applied to the end caps. As the amplitude on the ring electrode increases, the frequencies of the ion oscillations also increase. When the resonant frequency of an ion reaches the end cap frequency, the ion will become excited into an oscillating motion that is so large that it becomes destabilised and is ejected from the trap along the axis of the end caps. Since the oscillating frequencies of the ions are a function of their mass, ions of different *m/z* values will exit the ion trap at different voltages and, therefore, times.

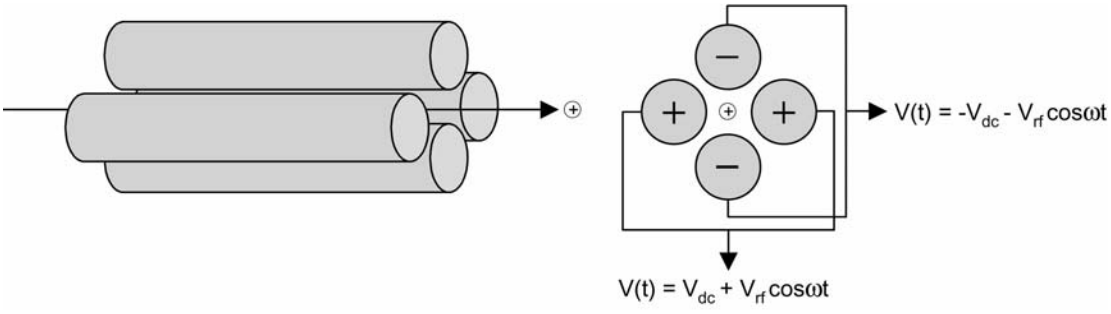

Figure 6. Schematic diagram of a quadrupole mass analyser, showing the direction of ion travel and the equations for the potentials applied to the rods. V(t) = voltage at time t; V_{dc} = direct potential; $V_{rf}\cos\omega t$ = radio frequency component. Adapted from Willoughby et al. [19].

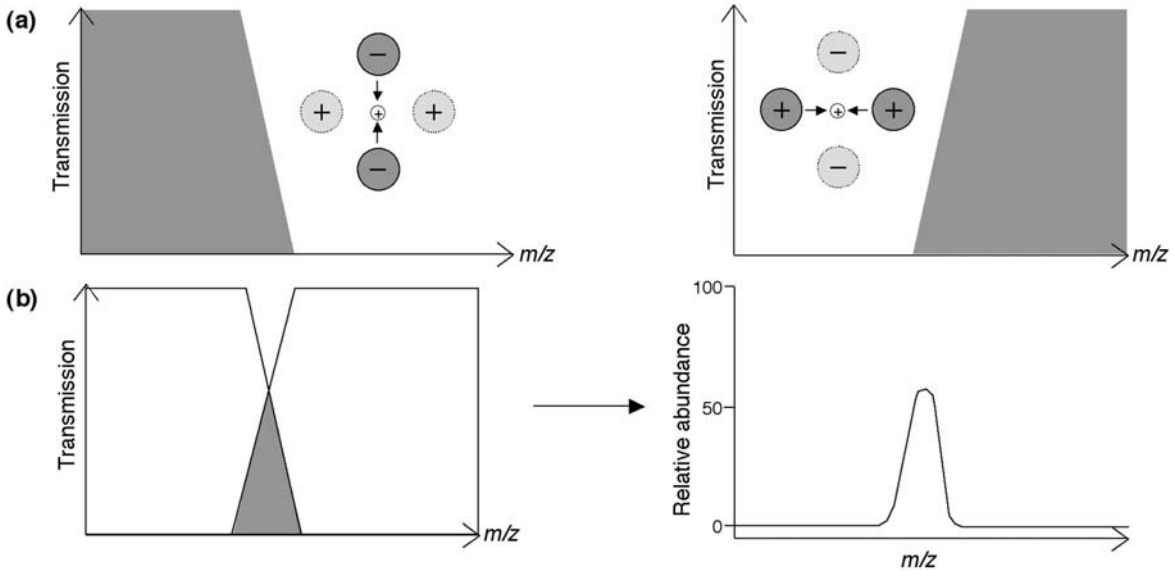

Figure 7. Mass filter created by a quadrupole mass analyser. (a) Negative rods create a low pass mass filter (left); positive rods create a high-pass mass filter (right). (b) Combining negative and positive sets of rods into a quadrupole arrangement overlaps the two mass filter regions, creating an area of mutual stability which allows ions of a certain *m/z* to pass. Adapted from Willoughby et al. [19].

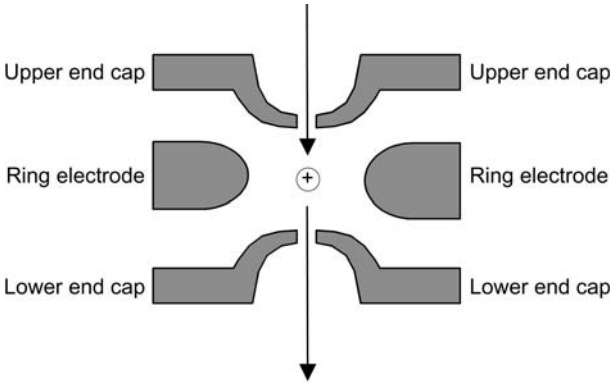

Figure 8. Schematic diagram of an ion trap mass analyser. Adapted from Willoughby et al. [19].

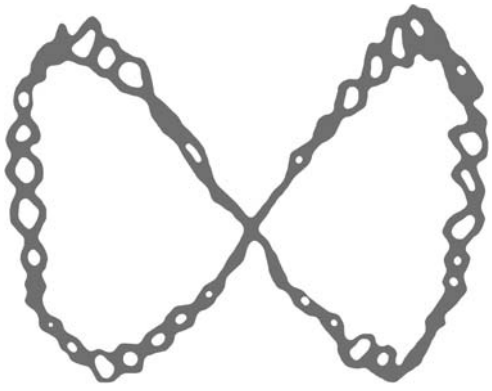

Figure 9. Oscillating figure of eight-shaped trajectory adopted by ions in an ion trap.

MS^n can be carried out in ion traps. Ions are selected for MS/MS analysis by using an rf voltage applied to the end caps to selectively eject all ions in the trap except for a chosen precursor ion. A resonating frequency that corresponds to the resonant frequency of the isolated ion is then applied to the end caps at an amplitude of a few percent of that required to eject the ion. The precursor ion starts to oscillate and collide with the helium buffer gas, which eventually induces fragmentation of the precursor ion (CAD). The fragment ions have different resonating frequencies from the parent ion and therefore are dampened into the centre of the trap by the helium gas. After a period of time, the rf voltage on the ring electrode is ramped, causing ejection of the ions in the manner described above. For MS^n analysis, this cycle is repeated to allow sequential stabilisation and subsequent fragmentation of successive product ions.

Ion traps are robust, sensitive and relatively inexpensive, and so are widely used for proteomics research. A disadvantage of the ion trap is its relatively low mass accuracy, due partly to the limited number of ions that can be accumulated in its centre before space-charging distorts their distribution, and therefore the accuracy of the measurement. Recently, the 'linear' ion trap has been developed [24], which stores the ions in a cylindrical volume that is considerably larger than that of traditional ion traps, allowing increased sensitivity, resolution and mass accuracy.

Time of flight

Analysis by TOF is based on the following principles: an accelerating potential (V) will give an ion of charge z an energy of zV, which can be equated to the kinetic energy of the ion:

$$zV = \frac{mv^2}{2}$$

where m = mass, v = velocity

If all ions are accelerated with the same potential, ions of different mass with the same charge must be travelling at different velocities. But velocity (v) = distance (d)/time (t), and therefore the equation can be rewritten:

$$\frac{m}{z} = \frac{2Vt^2}{d^2}$$

and ions with different mass will take different amounts of time to travel the same distance.

Mass-to-charge ratios are determined by measuring the time that ions take to move through a field-free region between the source and the detector. Mass resolution is affected by slight variations in flight time, and factors that create a distribution in flight times among ions with the same m/z ratio will result in poor mass resolution. Two techniques are used to compensate for temporal (time of ion formation), spatial (location of ion formation) and kinetic (energy of ion formation) distribution. By introducing a time delay between ion formation and extraction of ions from the source, wide spatial and temporal distributions can be avoided. This is known as 'delayed extraction'. Ions are first allowed to expand into a field-free region in the source, then after a certain delay (nanoseconds to microseconds), a voltage pulse is applied to extract the ions outside the source. The second technique is the use of ion mirrors, or reflectrons, which compensate for variations in energy distribution. The reflectron creates a retarding field that deflects the ions, sending them back through the flight tube. The more energetic the ion, the deeper it penetrates the retarding field of the reflectron before being reflected. Thus a more energetic ion will travel a longer flight path and arrive at the detector at the same time as less energetic ions of the same mass. A schematic diagram of a reflectron TOF analyser is shown in figure 10.

MS/MS can be achieved with reflectron TOF analysers by the observation of post-source decay (PSD) fragments; however, this approach is protracted and arduous. MS/MS with TOF analysis is most commonly practiced by placing a collision cell between two TOF analysers, or by configuring a TOF analyser as the second stage in hybrid instruments (see below).

Fourier transform MS

A Fourier transform mass spectrometer (FTMS), also known as an ion-cyclotron resonance mass spectrometer (ICR-MS), is a type of ion trap consisting of a cubic cell inside a strong magnetic field. The cell has three sets of plates: trapping, transmitter and receiver (fig. 11). For mass analysis, a continuous beam of ions is formed outside the cell and guided up to the trap, and a portion of the beam is pulsed into the cell. Inside the cell the ions are constrained by the strong magnetic field so that they move in cyclotron motion (circular orbit, see fig. 11) in a plane perpendicular to the magnetic field; they are also constrained by electric potentials applied to the trapping plates, which are also perpendicular to the magnetic field (fig. 11). Cyclotron motion is periodic and characterised by its cyclotron frequency, the frequency with which an ion repeats its orbit. The cyclotron frequency, f_c, is determined by the strength of the magnetic field (B), the charge on the ion (z) and the mass of the ion (m):

$$f_c = \frac{zB}{2\pi m}$$

Thus for a constant magnetic field, the m/z ratio of an ion is determined by measuring its cyclotron frequency. The radius of the cyclotron orbit depends on the kinetic energy of the ion and, since the cyclotron frequency is constant for an ion of given m/z, scales directly with the ve-

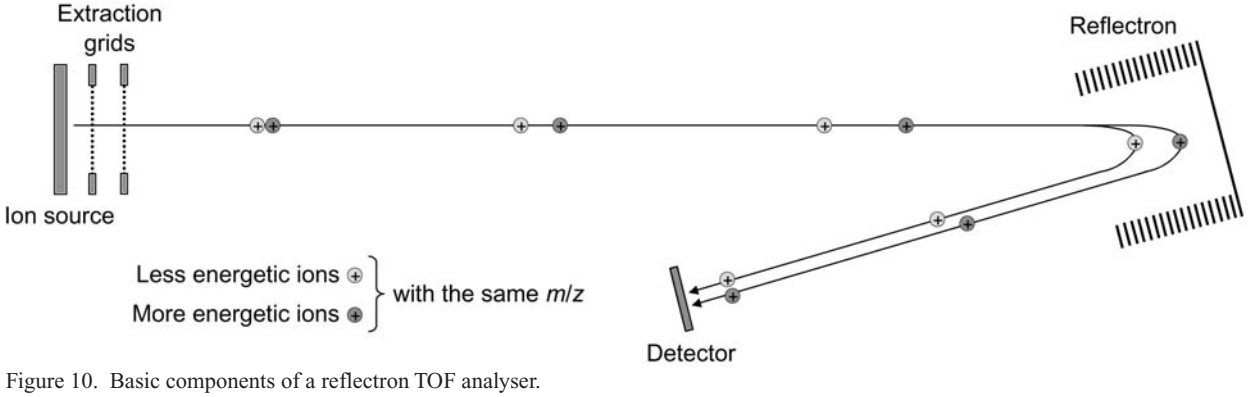

Figure 10. Basic components of a reflectron TOF analyser.

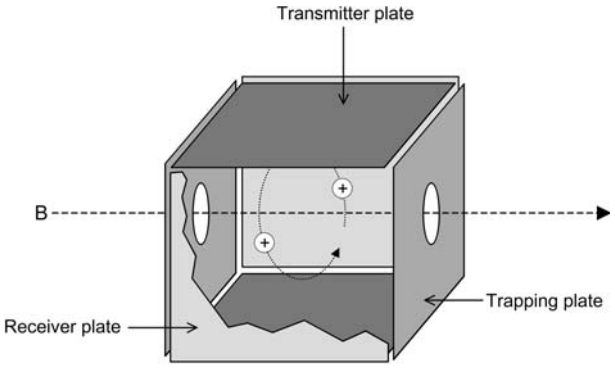

Figure 11. Schematic diagram of FTMS analyser. Ions are injected into the trap along the same plane as the magnetic field, B, and trapped by a voltage applied to the trapping plates. The ions move in cyclotron motion in the plane perpendicular to the plane of the magnetic field. Adapted from Willoughby et al. [19].

locity of the ion, or with the square root of the kinetic energy. When an ion is first trapped, the radius of its cyclotron orbit is usually small compared with the dimensions of the cell. To detect the ions, rf electric potentials are applied to the transmitter plates. An ion whose cyclotron frequency is in resonance with the frequency of the applied rf field will absorb energy and, as it does so, the radius of its cyclotron orbit will increase. As the excited ions pass the receiver plates, the frequency of their passage is detected as an induced current called the 'image' current. To accomplish the simultaneous detection of ions with many different *m/z* ratios, many frequencies are applied during the excitation event, most commonly by using a rapid-frequency sweep in a short (1 ms) time period [25]. The image current that results is a composition of the frequencies and amplitudes of ions of many different *m/z* values. The signal is converted to a mass spectrum by applying an FT (fig. 12). Once the ions have been detected, an rf 'quench' pulse is applied to eject the ions before the next lot of ions are introduced into the cell.

FTMS is unlike any other form of mass analyser because ion detection is non-destructive; signal-to-noise can be improved by averaging many cycles before transforming and storing the data.

MS/MS analysis with FTMS is similar to that in the ion trap: ions are first isolated by ejecting all other ions in the cell using resonance excitation, then a pulse of gas is introduced into the cell and a small voltage is applied to the transmitter plates. By varying the amplitude of the resonating frequency, the precursor ion starts to oscillate and collide with the background gas, inducing dissociation and production of fragment ions. After a period of time, the ions are excited into higher cyclotron orbits and detected.

FTMS provides high sensitivity, mass accuracy, resolution and dynamic range, but instruments are expensive and their operation complex; to date, these two factors have limited their routine use in proteomics research, despite their substantial potential [26].

Coupling ion sources to mass analysers

MALDI is traditionally coupled to TOF analysers, which measure the mass of intact peptides, whilst ES is mostly coupled to ion traps and triple quadrupoles, which allow the generation of fragment ion spectra from selected precursor ions. In the last few years, MALDI ion sources have been coupled to quadrupole ion trap mass spectrometers [27], TOF-TOF analysers [28], in which two TOF sections are separated by a collision cell, and quadrupole-TOF (QTOF) analysers [29], in which a collision cell is placed between a quadrupole mass filter and a TOF analyser. These instruments have good sensitivity, resolution and mass accuracy, and allow the fragmentation of MALDI-generated precursor ions. In addition, the QTOF can be used interchangeably with an ES ion source.

Although ion trap, TOF and hybrid TOF instruments are currently the most widely used for proteomics research, FTMS and linear ion traps are likely to become widespread in the future.

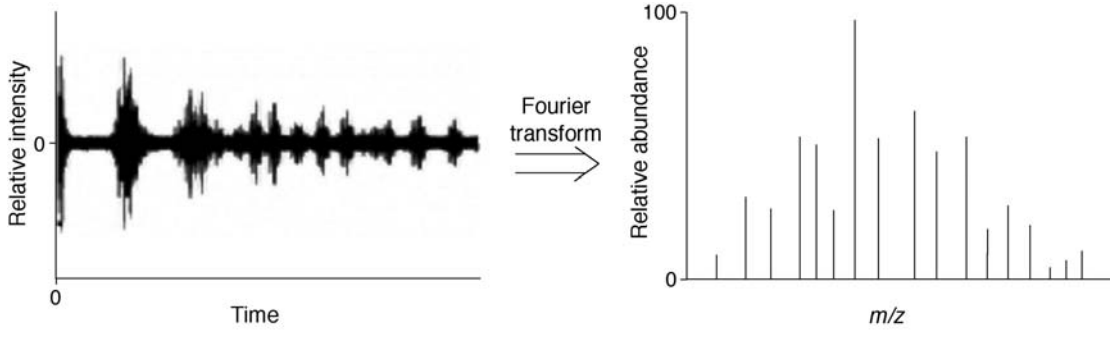

Figure 12. Time and mass domain for FTMS. Signal intensity as a function of time is transformed, through an FT, into signal intensity as a function of frequency, and hence to *m/z*.

Peptide fragmentation theory

Protein identification and peptide sequencing by MS/MS necessitates knowledge of how peptides fragment in the gas phase. The transfer of ions into the gas phase by ES is not a highly energetic process. The desolvation process effectively cools the ions, and therefore gaseous ions entering into the mass spectrometer for analysis have low internal energy [22]. The energy required for the ions to undergo dissociation is introduced by collisional activation. Peptide ion fragmentation is promoted by a 'mobile' proton, i.e. the fragmentation of most protonated peptides requires the involvement of a proton at the cleavage site [30]. The relative populations of the different protonated forms of a particular peptide depend on the internal energy content of the peptide and the gas phase basicities of the potential sites of protonation. The addition of energy via collisional activation alters the initial population of protonated forms (i.e. 'mobilises' the proton) and increases the population of protonated forms with energies higher than that of the most stable structure; for these excited molecules, the proton is located at various backbone heteroatoms. Protonation at backbone sites initiates cleavage of the backbone to produce b- and/or y-type fragments (fig. 13).

For singly protonated peptides containing a strongly basic amino acid, the energy required to relocate the proton to a position on the backbone and hence induce dissociation is considerably higher than for peptides containing no basic amino acids. Dissociation energy requirements are greatest for arginine-containing peptides and decrease in the order Arg-containing > Lys-containing > non-basic, i.e. in order of decreasing gas phase basicity [31]. In doubly charged tryptic peptides, the initial sites of protonation are generally the basic side chain of the C-terminal residue and the primary amine group at the N terminus [22] (fig. 13). The energy barrier to transfer of the N-terminus proton to the peptide backbone and between different sites on the peptide backbone is low so that various different fragmentation pathways are promoted, resulting in MS/MS spectra with several different b and y ions. Tang and Boyd [32] observed that fragmentations of tryptic peptide $[M + H]^{2+}$ ions resulted in product ion spectra in which the y ions appeared at appreciably higher relative abundance than their b ion counterparts. This was explained by the higher stability of the y ions due to the sequestering of the proton on the basic side chain of the C-terminal residue, whereas the b ions have no strongly favoured site of charge, so are likely to undergo further intramolecular proton rearrangement and fragmentation. In addition, if high enough collisional activation is achieved, b ions can undergo further dissociation to form a ions, with concurrent loss of carbon monoxide (fig. 14).

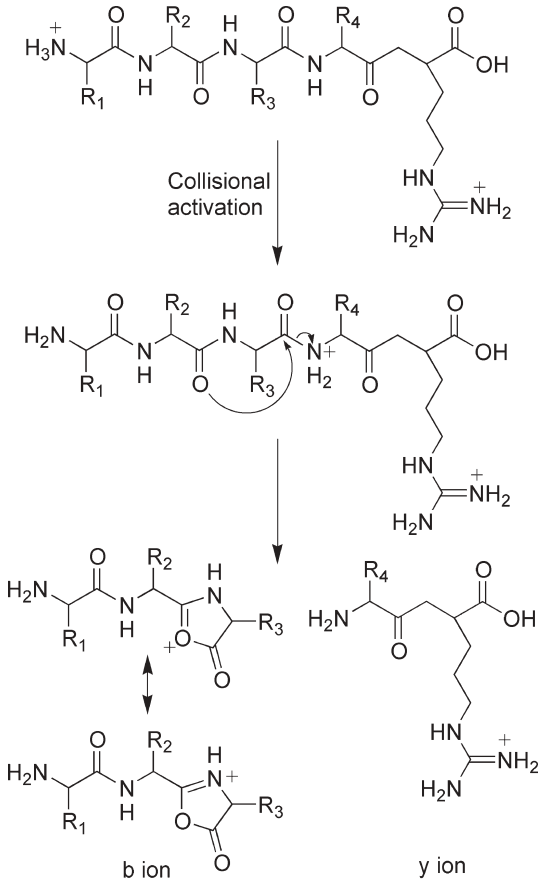

Figure 13. Fragmentation of a doubly charged tryptic peptide to produce b and y ions. Relocation of the N-terminus proton to promote charge-directed cleavage via CAD may occur to either N or O backbone atoms [30].

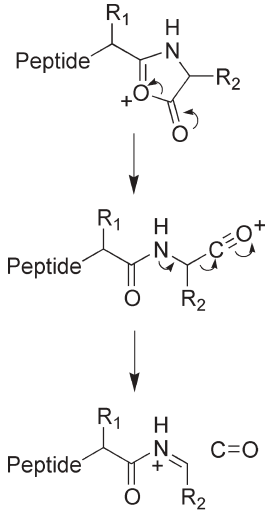

Figure 14. Dissociation of b ions to produce a ions and carbon monoxide.

Residue-specific cleavages

The presence of certain amino acid residues in a peptide will result in the selective cleavage of the peptide at those sites. Enhanced cleavage is observed N-terminal to proline residues because, as a tertiary amine, the backbone N atom of proline is more basic than the other backbone N atoms and the mobile proton is therefore more likely to reside at this position [33–35]. Conversely, fragmentation C-terminal to proline is rare because of steric hindrance [36] (fig. 15).

For peptides containing arginine and aspartic and/or glutamic acid, enhanced cleavage occurs C-terminal to the acid residue if the number of added protons is equal to or less than the number of arginines present [30]. In these cases, the arginine(s) tightly binds the proton(s) allowing the acidic hydrogen of the acid side chain to initiate cleavage (fig. 16).

Where the number of added protons exceeds the number of arginines present, enhanced cleavage at acidic residues is not observed. For peptides containing lysine rather than arginine and aspartic acid and/or glutamic acid there is no enhancement of cleavage at acidic residues [30].

In contrast to the results for acidic residues, where selective cleavage occurs if there is no added proton available to catalyse fragmentation, enhanced cleavage occurs C-terminal to histidine when an added proton is present on the histidine side chain [30]. All singly protonated peptide ions containing both histidine and arginine fragment non-selectively, yet doubly protonated peptide ions with arginine and histidine, or singly protonated peptides containing histidine but not arginine, show selective cleavage. This is explained mechanistically in figure 17.

Depending on the identity of the amino acid residue C-terminal to histidine (R_2, fig. 17), proton transfer may occur from the histidine ring to the C terminal-leaving fragment.

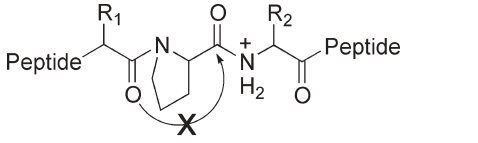

Figure 15. Cleavage C-terminal to proline residues is rare due to steric hindrance.

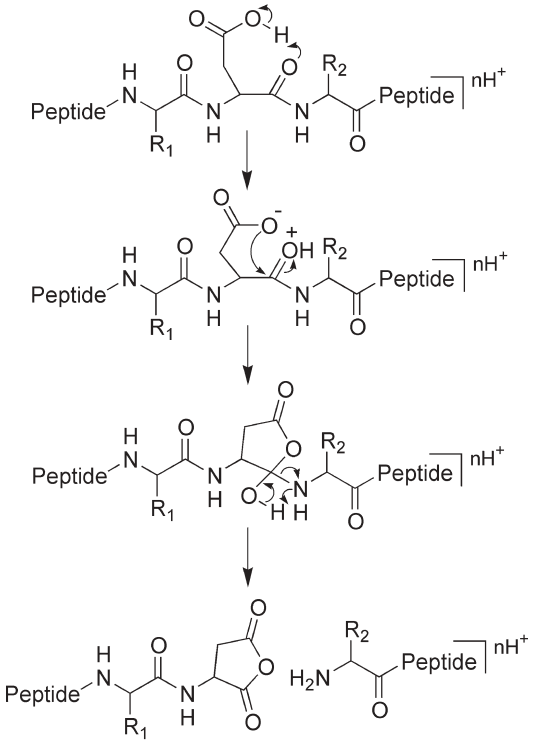

Figure 16. Selective cleavage occurs at positions C-terminal to acidic residues when the number of added protons (n) is equal to or less than the number of arginines present.

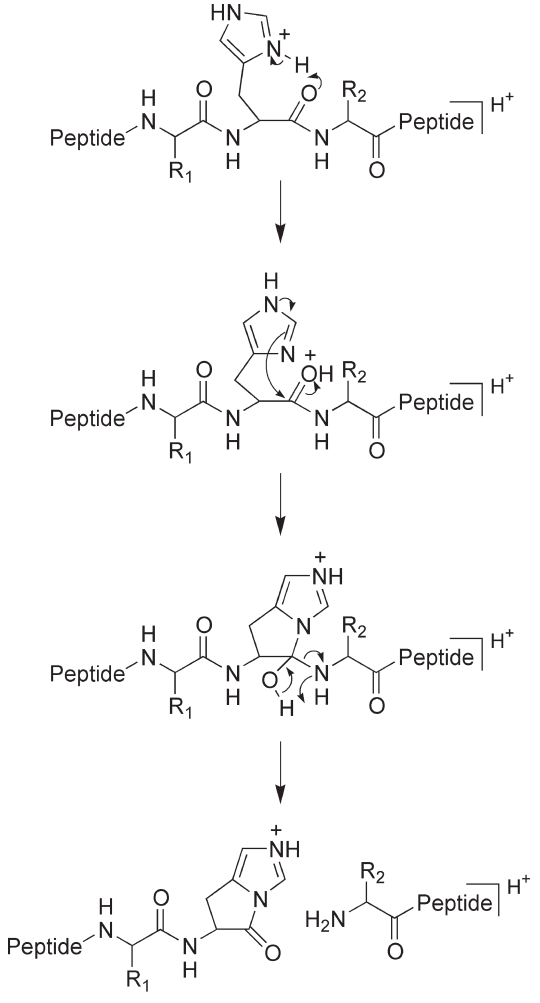

Figure 17. Enhanced cleavage is observed C-terminal to histidine when an added proton is present on the histidine side chain [30].

Protein identification

There are two main routes by which proteins are identified using MS. The classic proteomics approach involves the separation of the proteins in a mixture by two dimensional gel electrophoresis (2DE), followed by in-gel tryptic digestion and peptide mass fingerprinting (PMF) by MALDI-TOF MS. In this approach, proteins are identified by matching the list of observed peptide masses with a calculated list of all the expected peptide masses for each entry in a protein database. Observed 2DE-separated protein spots can be quantified by staining intensity. However, there are many problems with 2DE. It has only a very limited dynamic range: only the most abundant proteins are observed [37], and several classes of proteins are known to be excluded or under-represented in 2DE patterns; these include very acidic or basic proteins, very large or small proteins and membrane proteins [38]. Improvements in 2DE technology such as sample prefractionation [39, 40], large-format, higher resolving gels [41], narrow pH-range gels [40] and more sensitive staining methods [42, 43] have alleviated, but not eliminated these limitations.

PMF requires relatively pure samples, since mixtures of proteins will create mixtures of PMFs, making protein identification difficult [44]. MALDI-MS/MS (see above) should facilitate the use of MALDI for the analysis of more complex samples, thus uncoupling MALDI-MS from 2DE. However, some prior separation of the sample will still be necessary; if LC is used, the effluent from the chromatographic run must be deposited on a sample plate and mixed with MALDI matrix. Several systems have been designed to facilitate this [45–48], but only now is automation of the process being implemented [26].

The second major approach to identifying proteins and peptides by MS is the use of LC-ES-MS/MS. This method is at the centre of MS-based proteomics today; a generic LC-ES-MS/MS approach is illustrated in figure 18.

One dimension of peptide chromatography does not provide sufficient peak capacity to separate peptides generated from complex protein mixtures; to address this, various different combinations of protein and peptide separation schemes have been explored involving two- or three-dimensional chromatography and/or one-dimensional gel electrophoresis (1DE). A popular method is the 'multi-dimensional protein identification technology', or MudPIT approach [49, 50], which makes use of the orthogonal separation methods of strong cation exchange (SCX) followed by reverse-phase (RP) chromatography to separate peptides generated from proteolytic digestion of entire protein mixtures. The number of proteins identified using the MudPIT approach can far exceed the number identified using 2DE experiments, for example 1484 unique proteins were identified from 5540 unique peptides from a total cell lysate of *Saccharomyces cerevisiae*, compared with around 300 proteins identified after 2DE experiments using similar samples [50].

A major stumbling block in the use of LC-ES-MS/MS to identify peptides from biological matrices is that of informatics: the amount of data generated by the method is huge, and its analysis can be extremely daunting. This is discussed in more detail below.

LC-ES-MS/MS and the bioinformatics bottleneck

Protein identifications from peptide MS/MS spectra are less ambiguous than those from PMF because the peak pattern in the MS/MS spectrum provides information additional to the measured mass of the peptide. However, the amount of data generated is huge: many thousands of peptide CAD spectra can be acquired during an LC-ES-MS/MS run. To identify these peptides, spectra are scanned against protein sequence databases using a search algorithm. Several algorithms have been developed for this purpose [51]. The most commonly used algorithms are Sequest [52], Mascot [53] and MS-Tag [44, 54]. Sequest adopts a cross-correlation approach, in which peptide amino acid sequences from a protein database are used to construct theoretical mass spectra, and the degree of overlap, or cross-correlation, between the theoretical and experimental mass spectra determines the best match. The Mascot method employs probability-based matching: the MS/MS fragment masses calculated from peptide sequences in the database are compared with the experimentally observed peaks, and a score is calculated that reflects the statistical significance of the match between the experimental and theoretical spectra. The MS-Tag approach involves extracting a short, unambiguous section of the amino acid sequence of the peptide from the experimental data, which is used along with the measured mass of the peptide to determine the protein of origin. For all of these approaches, identified peptides are compiled into a protein 'hit-list'. Currently available de novo sequencing algorithms are computationally intensive and require high-quality data [51]. Therefore, for high-throughput proteomics studies, protein identifications are restricted to those proteins whose sequences appear in the searched database; many post-translationally modified or mutated proteins will be overlooked.

The algorithms used for searching MS/MS data are not infallible, and the main challenge in the interpretation of database search results is how to distinguish false-positive identifications. If the best matches in all database searches were assumed to correspond to the correct peptides, then a large proportion of these assigned peptides would be wrong [55]. This situation can arise for several reasons: the scoring schemes used in current database

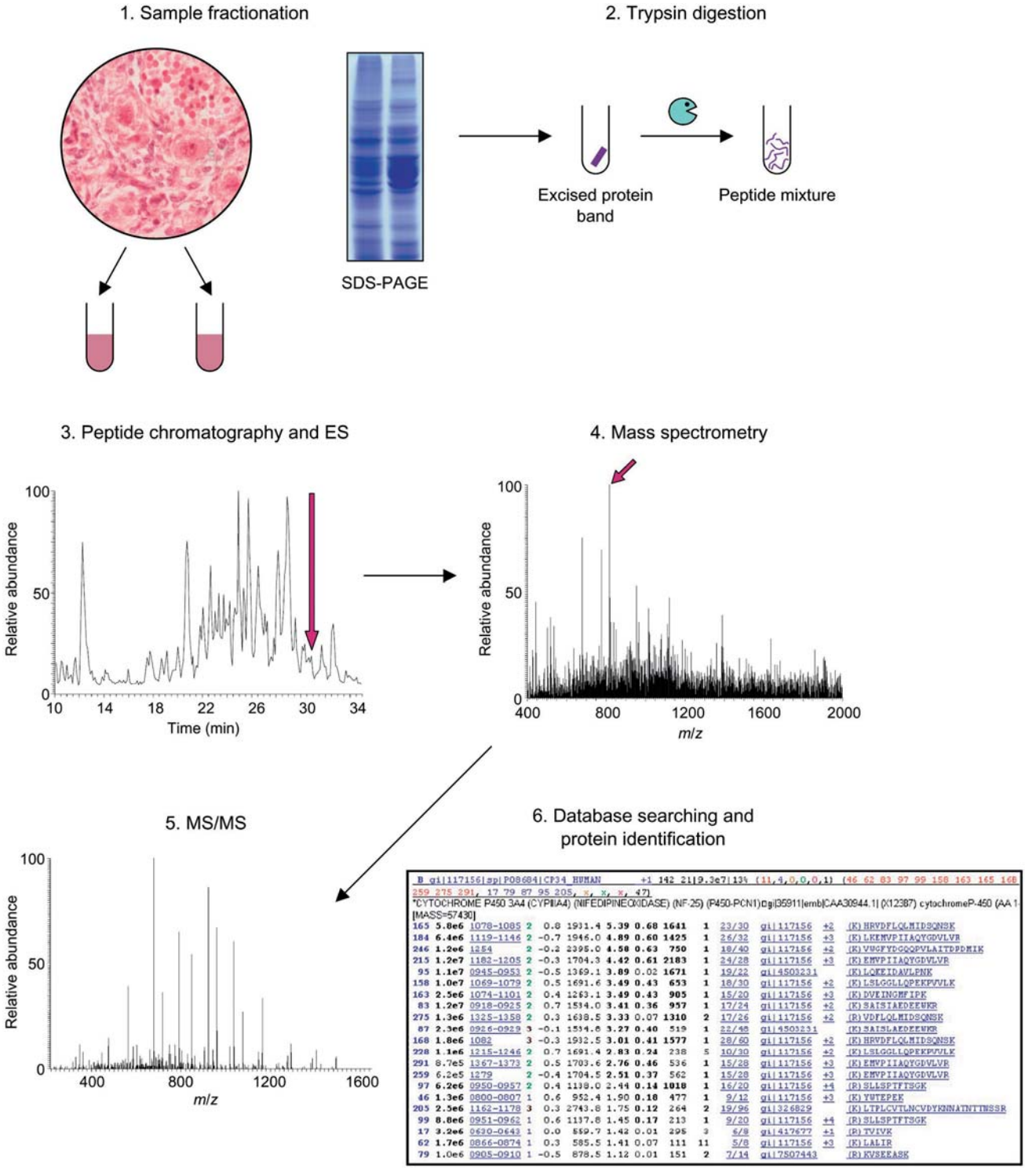

Figure 18. A generic MS-based proteomics experiment, consisting of six stages. Proteins for analysis are first isolated from cell lysate or tissues by fractionation methods, often including a final step of SDS-PAGE. Proteins are then enzymatically digested and separated using one or more stages of LC. The LC eluate is directed into an ES source, and the peptides are analysed first by full-scan MS and then by MS/MS. Peptide MS/MS spectra are searched against a protein database using a search algorithm which assigns peptide identifications based on match criteria.

search tools are based on a simplified representation of the peptide ion fragmentation process [36, 56]; the charge states of the peptides selected for fragmentation are not always known; many MS/MS spectra are of poor quality or are mixtures of the fragmentations of more than one peptide; high-quality MS/MS spectra will be incorrectly assigned if their true corresponding peptides are not in the database. Manual verification of peptide assignments is time consuming and is not feasible for the analysis of data sets that may contain tens of thousands of spectra.

The number of incorrect peptide identifications can be reduced by additional processing of the data before searching, by, for example, removal of low-quality spectra, clustering of redundant spectra and application of charge state determination algorithms [57–62]. Additionally, the introduction of more advanced scoring systems that include additional knowledge of how peptides fragment [36, 56, 63] should generate further improvements.

To attempt to separate correct from incorrect peptide assignments, filtering criteria, based on database search scores and other available data, can be applied. However, different researchers often use their own preferred filtering criteria, making it difficult to compare results between (or even within) research groups. Software tools such as DTASelect [64], INTERACT [65] and CHOMPER [66], compatible with Sequest and Mascot, are available to facilitate the filtering of data. Yet the rates of false identifications resulting from the use of these filters are rarely estimated [51]. Consequently, the question of what constitutes protein identification in an LC-MS/MS experiment is difficult to answer.

To allow the comparison of proteomic data between groups, peptide assignments should be validated using statistical programs developed to be compatible with existing database search tools [51]. The development and application of robust, transparent tools for the statistical analysis of proteomic data is essential. Several computational methods have recently become available [67–72]. Only when these programs have become standardised and widely used can one of the long-term aims of proteomics begin to be realised: the creation of centralised, public databases of peptide and protein identifications with repositories for storing MS data. The combined results from many different research groups could then be merged and applied to the whole genome, validating expressed genes at the protein level, and enabling the elucidation of patterns of protein expression that would be missed in individual experiments.

Protein quantification

In a recent article in *Nature* [26], Aebersold and Mann argue that for the documentation of the expression of proteins as a function of cell or tissue state to be meaningful, the data must be at least semi-quantitative: a list of the proteins detected in the different states is inadequate.

In both MALDI- and ES-MS, the relationship between the concentration of an analyte in a sample and its measured signal intensity is dependent on multiple factors that are difficult to control and are, in fact, not completely understood [26]. Consequently, in the absence of internal standards, mass spectrometers are inherently poor quantitative devices. Quantitation in MS-based proteomics is most commonly achieved by using stable isotope dilution. This makes use of the fact that two analytes differing only in stable isotope composition will be chemically almost identical, eluting at very similar retention times, yet can be differentiated in a mass spectrometer due to their mass difference. The ratio of the ion currents for the analyte pair will be equal to the abundance ratio for the two analytes. Therefore, relative abundances of proteins in different cell or tissue states can be calculated, and absolute abundances can be measured with the use of isotopically labelled standards. The three main ways of introducing stable isotopes to proteins are via chemical reactions, by metabolic labelling using heavy isotope-labelled salts or amino acids, and via enzymatic incorporation of ^{18}O from $H_2{}^{18}O$ during proteolysis.

The most widely used chemical labelling method is the isotope-coded affinity tag (ICAT) approach [73]. The ICAT reagent contains a biotin affinity tag, a linker that can incorporate stable isotopes and a thiol-specific reactive group. The reagent exists in heavy and light forms. Proteins from two different cell states are harvested, denatured, reduced and labelled at cysteine residues using the light or heavy ICAT reagents. The samples are then combined and digested with trypsin. The ICAT-labelled peptides are isolated using the biotin affinity tag by affinity chromatography and analysed by LC-ES-MS/MS. The ratio of the ion intensities for ICAT-labelled pairs of peptides allows the calculation of the relative abundance of the parent protein in the two cell states, whilst MS/MS confirms the sequence of the peptide and identifies the protein. The ICAT strategy is shown in figure 19. However, the original ICAT reagents were relatively large, and the presence of their fragments in the MS/MS patterns of peptides was found to complicate the analysis. To address this, a cleavable ICAT reagent was introduced [74], which contains an acid-cleavable linker that allows the removal of the affinity tag before MS analysis of the peptides. In addition, ^{2}H-labelling was replaced by ^{13}C-labelling, to eliminate chromatographic displacement effects. The disadvantage of the ICAT approach is that it is limited to the analysis of cysteine-containing proteins; however, this feature can be utilised to simplify complex mixtures of proteins.

One of the features of the chemical-tagging approach is that isotope-tagging reactions can be chosen to direct the isotopes and attached affinity tags to specific functional

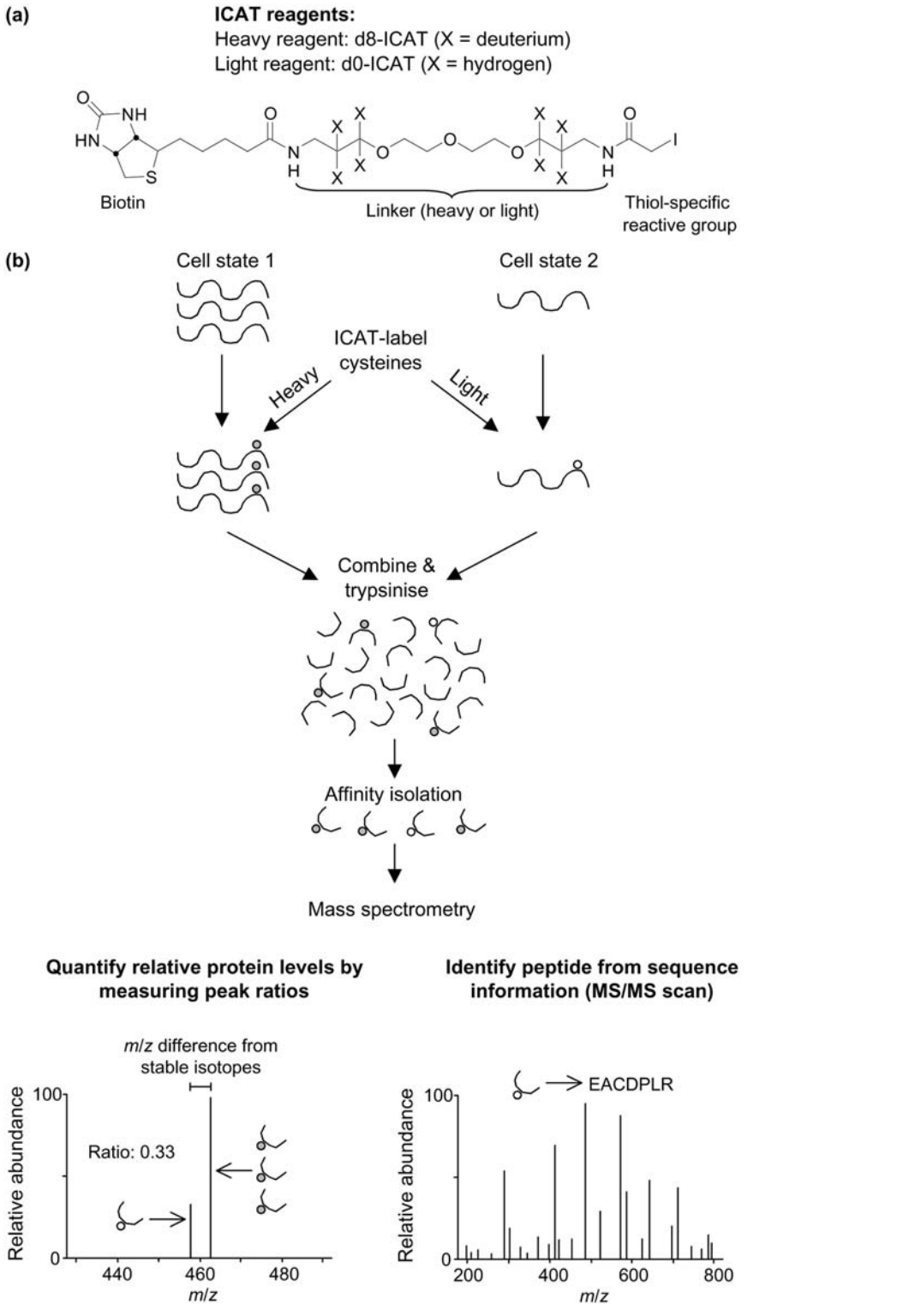

Figure 19. The ICAT strategy for quantifying differential protein expression. Adapted from Gygi et al. [38]. (a) Structure of the original ICAT reagent. (b) Schematic of the ICAT strategy.

groups or protein classes, allowing selective isolation and analysis of the tagged peptides. Tagging reactions that are specific for sulphydryl groups [73, 75], amino groups [76], the active sites of serine [77] and cysteine hydroxylases [78], phosphate ester groups [79, 80] and N-linked carbohydrates [81] have been described.

Metabolic isotope labelling was, until recently, limited to the study of single-celled organisms or cells grown in culture. The 'stable isotope labelling with amino acids in cell culture', or SILAC, approach [82] involves growing mammalian cell lines in medium lacking a standard essential amino acid but supplemented with an isotopically labelled form of that amino acid, for example lysine deuterated at its side chain methylenes. Potentially, all peptides can be labelled [82]. In other studies [83–86], cells were grown in medium enriched with stable isotopes such as ^{15}N. Krijgsveld et al. [87] have taken this approach one step further by metabolically labelling the nematode *Caenorhabditis elegans* and the common fruit fly *Drosophilia melanogaster* by feeding them on ^{15}N-labelled *Escherichia coli* and yeast, respectively. After two generations, ^{15}N incorporation was 98% for the nemotodes and 94–95% for the fruit flies.

When samples are of human origin, labbelling proteins metabolically is not practical. Enzyme catalysed incorporation of ^{18}O from $H_2{}^{18}O$ during proteolysis [88–91] results in each peptide generated being labelled at its carboxy terminal. Analysis is therefore not limited to specific functional groups or protein classes, in contrast to chemical-tagging approaches. The mass difference generated by ^{18}O incorporation is 4 Da, which can make quantitation difficult because isotope patterns for analyte pairs are likely to overlap. However, this methodology has been successfully applied to complex biological samples [91].

A simple approach for absolute quantitation has recently been described by Gerber et al. [92]. This involves the use of synthetic peptides, specific to the protein of interest, with incorporated stable isotopes. Samples are separated by 1DE and the bands of interest subjected to in-gel tryptic digestion in the presence of known amounts of the labelled peptide(s). Absolute quantitation is then achieved using LC-MS/MS analysis. There is also the potential for whole proteins to be expressed using only ^{15}N- or ^{13}C-containing amino acids; these proteins could be introduced to samples prior to SDS-PAGE separation, thereby eliminating any differences between sample and standard due to incomplete trypsinisation.

Stable-isotope dilution methods in combination with LC-ES-MS/MS are increasingly being used to detect changes in protein profiles, for example between diseased and normal tissue states, and to infer biological function from the observed patterns.

Protein interactions

The majority of proteins exert their function via protein-protein interactions. To determine the binding partners of a protein by MS, the protein itself can be used as an affinity reagent. There are three crucial components to these MS-based protein interaction experiments: bait presentation, affinity purification of the complex and analysis of the bound proteins [26]. If antibodies to the protein of interest are available, these can be used to specifically isolate the protein with its bound partners. Alternatively, the protein of interest can be 'tagged' with a sequence recognised by an antibody specific for the tag. The tagged protein can be expressed in stable cell lines and allowed to bind to any other proteins with which it interacts. The protein and its binding partners can be separated from the mixture using the antibody, with subsequent denaturing of the entire complex for analysis by MS. The tandem affinity purification (TAP) approach [93, 94] combines the use of two different tags on the same protein, separated by an enzyme-cleavable linker sequence. Proteins are tagged with a calmodulin-binding peptide linked to a tobacco etch virus (TEV) protease site linked to a protein A domain (fig. 20). The protein A domain binds tightly to a solid support modified with immunoglobulins. After washing, treatment with TEV protease allows the elution of the bound material, which is then incubated with calmodulin beads in the presence of calcium. This second affinity step allows the removal of any remaining contaminants as well as the TEV protease. The bound material is released with ethylene glycol-bis(2-aminoethylether)-N,N,N′,N′-tetraacetic acid (EGTA). The use of TAP tags reduces background noise considerably compared with tags supporting single-step purification, but may result in the loss of some of the weaker and more transient binding partners in the second affinity step [26].

Introducing a tagged protein into a cell system can cause problems because tagged protein expression levels are usually different from the expression levels of the untagged endogenous counterparts: artefacts can arise due to non-physiological levels of the tagged protein [26]. Also, many protein-protein interactions are transient, of low affinity and dependent on a specific cellular environment, therefore MS-based experiments of the type described above will only detect a subset of the interactions that actually occur. This is illustrated by a comparison of the results obtained in two different studies [95]. Gavin et al. [96] introduced 1739 TAP-tagged genes into the yeast genome; 232 stable complexes were isolated and the proteins identified by MALDI PMF after 1DE separation. In a similar study, Ho et al. [97] expressed Flag epitope-tagged proteins; single-step immunopurification was used to isolate the complexes, followed by LC-MS/MS for identification of proteins separated by 1DE. Both studies reported a large number of interacting proteins,

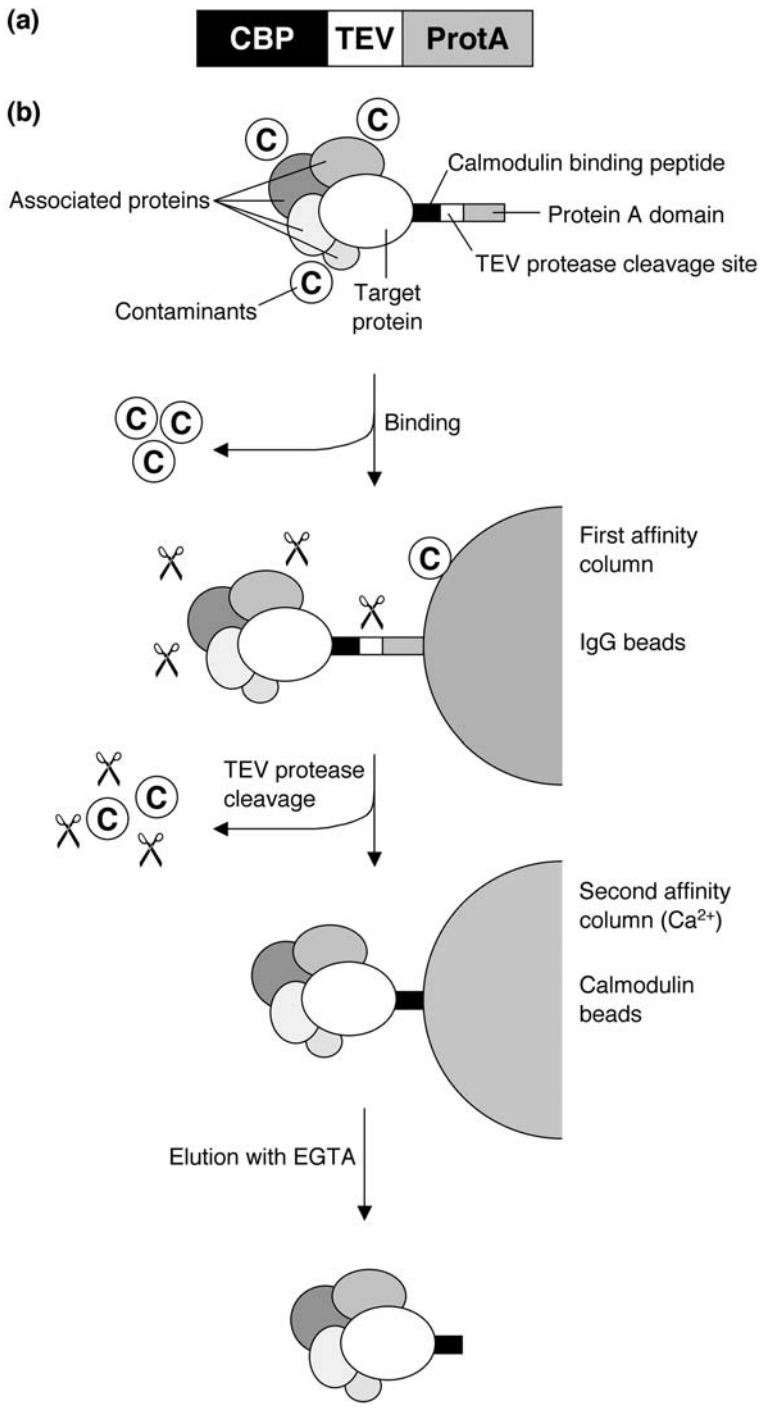

Figure 20. The TAP approach to the detection of protein interactions. Adapted from Puig et al. [94]. (a) Schematic representation of the TAP tag. (b) Overview of the TAP purification strategy.

but, interestingly, for the group of bait proteins that was common to both studies, surprisingly little overlap of data was observed, either between the two MS studies or between these and previous yeast two-hybrid studies [95]. Both MS studies reported results consistent with existing literature for known protein complexes. These results point towards the potentially huge scale of the 'interactome', and indicate that it is largely under-sampled in current studies.

In the future, stable-isotope methods are likely to revolutionise the study of protein-protein interactions. Stable-isotope ratios can be used to distinguish between the protein composition of two or more complexes, and between true complex components and non-specifically associ-

ated proteins [98, 99]. This ability to distinguish complex components from a background of non-specific binding will allow for fewer purification steps and less washing, thereby increasing the chances of detecting weak and transient interactions.

MS-based proteomics has also been used to study large protein complexes such as the yeast and human spliceosomes [100, 101], the yeast nuclear pore complex [102] and the human nucleolus [103].

Protein modifications

Post-translational modifications (PTMs) are chemical processing events that alter the properties of a protein after its translation, either by proteolytic cleavage or by addition of a modifying group to one or more amino acids. PTMs include phosphorylation, glycosylation, acetylation, methylation, sulphation, disulphide bond formation, deamidation and ubiquitination [104]. Most eukaryotic proteins are post-translationally modified, and many of these PTMs are regulatory and reversible, most notably protein phosphorylation, which is a dynamic process with complex kinetics involving several amino acids in a single protein, and which controls biological function through many different mechanisms [26].

The application of MS to the identification of PTMs ranges from the study of single, purified proteins through the search for one type of modification on all the proteins in a sample to scanning for all modifications on a proteome-wide scale. The complexities involved in identifying all the modifications even on a single protein mean that proteome-wide scanning is, at present, not comprehensive; nevertheless, large amounts of biologically useful information can be generated. To determine the sites of modification, maximum protein sequence coverage is desirable; for this purpose, peptide mass mapping using two or more different enzymes can be employed, for example trypsin, Asp-N and Glu-C. Protein modifications are then identified from the measured masses and fragmentation spectra using manual or computer-assisted interpretation [26]. However, the introduction of a number of possible modifications into computer search programs such as Sequest and Mascot can result in a 'combinatorial explosion' due to the need to consider all possible modifications for all peptides in the database [104]. To avoid this, the experiment can be divided into a two-stage process whereby a set of proteins is first identified on the basis of non-modified peptides, then only these proteins are searched for modifications; using this approach MacCoss et al. [105] identified 73 sites of modification on 11 different crystalline proteins from human lens tissue. Also available is SALSA (Scoring ALgorithm for Spectral Analysis) [106], which can be applied to MS/MS spectra that remain unassigned after initial database searching. SALSA enables the identification of modified peptides based on correspondence between theoretical and actual MS/MS ion series for peptide sequences, regardless of their absolute positions on the *m/z* axis.

Other approaches that attempt to address the low-stoichiometry and high-complexity problems associated with the analysis of PTMs, and protein phosphorylation in particular, involve the selective enrichment of modified proteins. These techniques are generally based on some form of affinity selection that is specific for the modification of interest, and that is used for the purification of modified proteins. Pandey and colleagues [107–109] immunoprecipitated tyrosine-phosphorylated proteins using antibodies specific for phosphotyrosine. Affinity purification has been combined with chemical modification: for example, Oda et al. [80] replaced phosphate moieties with affinity tags; Zhou et al. [79] used a sequence of chemical reactions to isolate phosphopeptides, and Goshe et al. [110, 111] employed a phosphoprotein isotope-coded affinity tag. Probably the most extensive characterisation of the phosphoproteome was achieved by Ficarro et al. [112], who esterified peptide mixtures to nullify negatively charged carboxyl groups, then captured the phosphopeptides by immobilised metal-affinity chromatography; more than 1000 phosphopeptides were detected from analysis of a whole-cell lysate from *S. cerevisiae*, and a total of 216 peptide sequences defining 383 sites of phosphorylation were determined. Peng et al. [113] used affinity purification to isolate the ubiquitinated proteins from yeast cells expressing 6xHis-tagged ubiquitin; 1075 proteins were identified, including 110 precise ubiquitination sites in 72 ubiquitin-protein conjugates. Two papers that identify N-glycosylated proteins have recently been published by Zhang et al. [81] and Kaji et al. [114]; both describe methods to immobilise the carbohydrate chain-containing proteins, then, after tryptic digestion, release the N-glycosylated peptides from their carbohydrate chains (using peptide-N-glycosidase F) for analysis.

Developing MS technologies are expected to substantially accelerate the analysis of PTMs. FTMS instruments can be coupled with electron capture dissociation (ECD), which has been shown to have profound potential for the characterisation of PTMs [115]. ECD occurs after recombination of multiply protonated protein or peptide molecules with thermal electrons. After electron capture, the resultant $[M + nH]^{(n-1)+\bullet}$ ion dissociates via energetic $H^{\bullet}$ transfer to the backbone carbonyl to form c (-CHR-C(OH)=NH) and $z^{\bullet}$ ($^{\bullet}$CHR′-) ions [116, 117]. Labile PTMs, which are easily lost under CAD conditions, remain intact during the ECD process, allowing the positions of modifications to be easily established; this is attributed to the high rate of bond cleavage and moderate amount of excess energy during ECD compared with the vibrational excitation involved in CAD [115]. FTMS and ECD have also been used for 'top-down' protein se-

quencing, in which the intact protein ion is fragmented inside the mass spectrometer, theoretically allowing the mapping of all PTMs [118, 119].

Future challenges and expectations

'The specific objective of proteomics is to concurrently identify, quantify and analyse a large number of proteins in a functional context' [26]. The global focus of analyses on a proteome-wide scale threatens to result in information overload, and throws up many challenges in terms of data collection, analysis and interpretation, visualisation, storage and data publication and sharing.

Proteomic studies generate huge amounts of data: in a typical LC-MS/MS experiment over 1000 CAD spectra of varying degrees of quality are generated in an hour. However, even if each one of these spectra led to the successful identification of a peptide, a very long time would be required to analyse complete proteomes. High-throughput collection of consistently high-quality data remains a challenge, and 'national proteome centres' have been suggested [120], to ensure availability of expertise and equipment.

The analysis and interpretation of the massive volumes of proteomic data are a major bottleneck in proteomics today. Expert manual analysis of data is incompatible with the thousands of spectra produced in a proteomics experiment; in addition, inconsistency exists between individuals. Therefore, the development of robust, transparent tools for the statistical analysis of proteomic data is crucial. Only when these tools have become standardised and widely accepted will the comparison of complementary proteomic data sets generated in different laboratories be possible. Such comparisons will also depend on well-organised and accessible systems for data storage, communication and visualisation; the development of these tools for proteomics is still in its infancy [26].

The current system for publication of scientific work does not lend itself well to the reporting of large data sets generated by proteomic experiments. Data tables are frequently published as supplementary to the main paper, and validation and discussion limited to only a few conclusions drawn from the data. To make the publication of proteomics data more useful, new ways must be devised to review and validate large data sets and to make their content electronically searchable. Preliminary developments by a few publishers and journals are described by Mann [121]; however, this problem remains essentially unsolved.

The sheer scale of proteomics research necessitates a community effort. Public access to online raw data would facilitate research in all areas, and enable those researchers with no access to MS data to lend their expertise to the development of more advanced computational methods and software tools. In a recent *Nature* commentary [122], Aebersold proposes a community-wide strategy involving the synthesis of one unique, heavy, stable isotope-tagged peptide marker for each protein, protein isoform or specifically modified form of a protein, which could then be used as a definitive marker for that protein. These markers would be added in precise amounts to samples where the proteins of interest had been labelled with natural-isotopic tags, and used for protein quantitation. This would drastically reduce the number of peptides that need to be analysed, and would trivialise the data analysis itself since de novo identification would no longer be necessary. The method could be easily standardised between laboratories, and the absolute quantitation would make different data sets easily comparable. Any subset of proteins could be interrogated, and, providing an appropriate reference peptide could be synthesised, any protein isoform, polymorphism or modified form could be analysed. Finally, once the initial cost of synthesising the reference peptides had been met, the method would be relatively cheap since only tiny amounts of the peptide standards would be needed for each assay.

MS-based proteomics is, however, still an emerging technology where radical change is possible, and for this reason the call for the standardisation of research approaches may be premature. Current instrumentation is far from optimal, partly because in the rapid explosion of proteomics research, manufacturers have not had the time to fully develop instruments perfectly customised for protein analysis. Mass spectrometers are still far from the physical limit of the few ions needed to register a peak, so a massive increase in performance can be expected in the next few years. There are several nascent techniques with the potential to revolutionise current MS-based proteomic practices. One of these, the analysis of intact proteins, has already been discussed. The emergence of whole-protein proteomics is being enabled by the development of mass spectrometers with large mass ranges, extremely high mass accuracy and resolution, and ionisation/fragmentation techniques compatible with large proteins; this has the potential to provide insights into modifications that are difficult to characterise by peptide analysis [26]. Another concept is the use of peptide 'accurate mass and time' (AMT) tags for high-throughput proteome characterisation [123–126]. This strategy involves the use of high mass accuracy and resolution FTMS to validate peptide AMT tags from 'potential mass tags' identified from global enzymatic digestion of a specific organism, tissue or cell type using conventional MS/MS techniques. Once the peptides have been validated they can be identified in subsequent experiments by simply correlating the accurate mass and the retention time with the database of previously determined AMT tags, thus obviating the need to perform MS/MS analyses. The advantage of using this method over MS/MS-

based techniques is the large number of protein identifications that can be obtained in a single LC run: the number of peptide identifications from an LC-MS/MS run is restricted because of the time needed to fragment the peptides individually. A third pioneering technique is that of imaging MS in which thin tissue sections are mounted onto a target plate, spotted with matrix and then analysed by MALDI [127]. From the intensity of a given *m/z* value, a density map or image is constructed; in this way, profiles of the proteins contained in the section can be obtained. The technique has already been applied to the assessment of protein patterns in several diseased tissues including human gliomas and non-small-cell lung cancer [128].

Outlook

MS is at the core of proteomics today. Its extensive range of existing applications encompasses the identification, characterisation and quantification of proteins, their interactions and modifications. Yet we are only at the tip of the iceberg: MS-based proteomics is still very much a nascent technology, where radical change is possible. Advances in instrumentation, experimental design and data handling will ensure that MS continues to play a pivotal role in life sciences research in the future.

Acknowledgements. C. S. L. is the holder of a C. W. Maplethorpe Postdoctoral Fellowship of the University of London. I would like to thank Dr. W. J. Griffiths and Prof. L. H. Patterson for helpful comments on the manuscript, and J. G. Lane for assistance with the figures.

1 Wilkins M. R., Pasquali C., Appel R. D., Ou K., Golaz O., Sanchez J. C. et al. (1996) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology **14:** 61–65
2 Tyers M. and Mann M. (2003) From genomics to proteomics. Nature **422:** 193–197
3 Pearson H. (2003) Geneticists play the numbers game in vain. Nature **423:** 576
4 Speicher D. W. (2003) Current challenges in proteomics: mining low abundance proteins and expanding protein profiling capacities. 16th International Mass Spectrometry Conference, Edinburgh, 31 August – 5 September
5 Thomson J. J. (1913) Rays of Positive Electricity and their Applications to Chemical Analysis, Longmans Green, London
6 Aston F. W. (1933) Mass Spectra and Isotopes, Arnold, London
7 Griffiths W. J., Jonsson A. P., Liu S., Rai D. K. and Wang Y. (2001) Electrospray and tandem mass spectrometry in biochemistry. Biochem. J. **355:** 545–561
8 Yamashita M. and Fenn J. B. (1984) Electrospray ion source: another variation on the free-jet theme. J. Phys. Chem. **88:** 4451–4459
9 Yamashita M. and Fenn J. B. (1984) Negative ion production with the electrospray ion source. J. Phys. Chem. **88:** 4671–4675
10 Meng C. K., Mann M. and Fenn J. B. (1988) Electrospray ionization of some polypeptides and small proteins. Proceedings of the 36th American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics, San Francisco, June
11 Fenn J. B., Mann M., Meng C. K., Wong S. F. and Whitehouse C. M. (1989) Electrospray ionization for mass spectrometry of large molecules. Science **246:** 64–71
12 Karas M. and Hillenkamp F. (1988) Ultraviolet laser desorption of ions above 10 kDa (abstracts) 11th International Mass Spectrometry Conference, Bordeaux, abstract MOA OR, 29 August – 2 September
13 Karas M. and Hillenkamp F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10000 Da. Anal. Chem. **60:** 2299–2301
14 Tanaka K., Waki H., Ido Y., Akita S., Yoshida Y. and Yoshida T. (1988) Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. **2:** 151–153
15 Klose J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues: a novel approach to testing for induced point mutations in mammals. Humangenetik **26:** 231–243
16 O'Farrell P. H. (1975) High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250:** 4007–4021
17 Dole M., Mach L. L., Hines R. L., Mobley R. C., Ferguson L. D. and Alice M. B. (1968) Molecular beams of macroions. J. Chem. Phys. **49:** 2240–2247
18 Gaskell S. J. (1997) Electrospray: principles and practice. J. Mass Spectrom. **32:** 677–688
19 Willoughby R., Sheehan E. and Mitrovich S. (1998) A Global View of LC/MS: How to Solve Your Most Challenging Analytical Problems, Global View, Pittsburgh
20 Thompson B. A. and Iribarne J. V. (1979) Field induced ion evaporation from liquid surfaces at atmospheric pressure. J. Chem. Phys. **71:** 4451–4463
21 Iribarne J. V. and Thompson B. A. (1976) On the evaporation of small ions from charged droplets. J. Chem. Phys. **64:** 2287–2294
22 Gaskell S. J. (1997) Electrospray: principles and practise. J. Mass Spectrom. **32:** 677–688
23 Hoffmann E. de and Stroobant V. (2003) Mass Spectrometry: Principles and Applications, Wiley, Chichester
24 Hager J. W. (2002) A new linear ion trap mass spectrometer. Rapid Commun. Mass Spectrom. **16:** 512–526
25 Amster I. J. (1996) Fourier transform mass spectrometry. J. Mass Spectrom. **31:** 1325–1337
26 Aebersold R. and Mann M. (2003) Mass spectrometry-based proteomics. Nature **422:** 198–207
27 Krutchinsky A. N., Kalkum M. and Chait B. T. (2001) Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. Anal. Chem. **73:** 5066–5077
28 Medzihradszky K. F., Campbell J. M., Baldwin M. A., Falick A. M., Juhasz P., Vestal M. L. et al. (2000) The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. Anal. Chem. **72:** 552–558
29 Loboda A. V., Krutchinsky A. N., Bromirski M., Ens W. and Standing K. G. (2000) A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. Rapid Commun. Mass Spectrom. **14:** 1047–1057
30 Wysocki V. H., Tsaprailis G, Smith L. L. and Breci L. A. (2000) Mobile and localized protons: a framework for understanding peptide dissociation. J. Mass Spectrom. **35:** 1399–1406
31 Dongre A. R., Jones J. L., Somogyi A. and Wysocki V. H. (1996) Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: evidence for the mobile proton model. J. Am. Chem. Soc. **118:** 8365–8374

32 Tang X. J. and Boyd R. K. (1992) An investigation of fragmentation mechanisms of doubly protonated tryptic peptides. Rapid Commun. Mass Spectrom. **6:** 651–657
33 Hunt D. F., Yates J. R. III, Shabanowitz J., Winston S. and Hauer C. R. (1986) Protein sequencing by tandem mass spectrometry. Proc. Natl. Acad. Sci. USA **83:** 6233–6237
34 Loo J. A., Edmonds C. G. and Smith R. D. (1993) Tandem mass spectrometry of very large molecules. 2. Dissociation of multiply charged proline-containing proteins from electrospray ionization. Anal. Chem. **65:** 425–438
35 Vaisar T. and Urban J. (1996) Probing the proline effect in CID of protonated peptides. J. Mass Spectrom. **31:** 1185–1187
36 Breci L. A., Tabb D. L., Yates J. R. III and Wysocki V. H. (2003) Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. Anal. Chem. **75:** 1963–1971
37 Gygi S. P., Corthals G. L., Zhang Y., Rochon Y. and Aebersold R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc. Natl. Acad. Sci. USA **97:** 9390–9395
38 Gygi S. P., Rist B. and Aebersold R. (2000) Measuring gene expression by quantitative proteome analysis. Curr. Opin. Biotechnol. **11:** 396–401
39 Naaby-Hansen S., Waterfield M. D. and Cramer R. (2001) Proteomics – post-genomic cartography to understand gene function. Trends Pharmacol. Sci. **22:** 376–384
40 Fey S. J. and Larsen P. M. (2001) 2D or nor 2D. Curr. Opin. Chem. Biol. **5:** 26–33
41 Gauss C., Kalkum M., Lowe M., Lehrach H. and Klose J. (1999) Analysis of the mouse proteome. I. Brain proteins: separation by two-dimensional electrophoresis and identification by mass spectrometry and genetic variation. Electrophoresis **20:** 575–600
42 Rabilloud T. (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics **2:** 3–10
43 Unlu M., Morgan M. E. and Minden J. S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis **18:** 2071–2077
44 Clauser K. R., Baker P. and Burlingame A. L. (1999) Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Anal. Chem. **71:** 2871–2882
45 Zhan Q., Gusev A. and Hercules D. M. (1999) A novel interface for on-line coupling of liquid capillary chromatography with matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun. Mass Spectrom. **13:** 2278–2283
46 Gelpi E. (2002) Interfaces for coupled liquid-phase separation/mass spectrometry techniques: an update on recent developments. J. Mass Spectrom. **37:** 241–253
47 Ericson C., Phung Q. T., Horn D. M., Peters E. C., Fitchett J. R., Ficarro S. B. et al. (2003) An automated noncontact deposition surface for liquid chromatography matrix-assisted laser desorption/ionization mass spectrometry. Anal. Chem. **75:** 2309–2315
48 Zhang B., McDonald C. and Li L. (2004) Combining liquid chromatography with MALDI mass spectrometry using a heated droplet interface. Anal. Chem. **76:** 992–1001
49 Washburn M. P., Wolters D. and Yates J. R. III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. **19:** 242–247
50 Wolters D. A., Washburn M. P. and Yates J. R. III (2001) An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. **73:** 5683–5690
51 Nesvizhskii A. I. and Aebersold R. (2004) Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. Drug Discov. Today **9:** 173–181
52 Eng J. K., McCormack A. L. and Yates J. R. III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. **5:** 976–989
53 Perkins D. N., Pappin D. J., Creasy D. M. and Cottrell J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20:** 3551–3567
54 Mann M. and Wilm M. S. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal. Chem. **66:** 4390–4399
55 Keller A., Purvine S., Nesvizhskii A. I., Stolyar S., Goodlett D. R. and Kolker E. (2002) Experimental protein mixture for validating tandem mass spectral analysis. OMICS **6:** 207–212
56 Kapp E. A., Schutz F., Reid G. E., Eddes J. S., Moritz R. L., O'Hair R. A. J. et al. (2003) Mining a tandem mass spectrometry database to determine the trends and global factors influencing peptide fragmentation. Anal. Chem. **75:** 6251– 6264
57 Gentzel M., Kocher T., Ponnusamy S. and Wilm M. (2003) Preprocessing of tandem mass spectrometric data to support automatic protein identification. Proteomics **3:** 1597–1610
58 Beer I., Barnea E., Ziv T. and Admon A. (2004) Improving large-scale proteomics by clustering of mass spectrometry data. Proteomics **4:** 950–960
59 Moore R. E., Young M. K. and Lee T. D. (2000) Method for screening peptide fragment ion mass spectra prior to database searching. J. Am. Soc. Mass Spectrom. **11:** 422–426
60 Kolker E., Purvine S., Galperin M. Y., Stolyar S., Goodlett D. R., Nesvizhskii A. I. et al. (2003) Initial proteome analysis of model microorganism *Haemophilus influenzae* strain Rd KW20. J. Bacteriol. **185:** 4593–4602
61 Sadygov R. G., Eng J., Durr E., Saraf A., McDonald H., MacCoss M. J. et al. (2002) Code developments to improve the efficiency of automated MS/MS spectra interpretation. J. Proteome Res. **1:** 211–215
62 Colinge J., Magnin J., Dessingy T., Giron M. and Masselot A. (2003) Improved peptide charge state assignment. Proteomics **3:** 1434–1440
63 Tabb D. L., Smith L. L., Breci L. A., Wysocki V. H., Lin D. and Yates J. R. III (2003) Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides. Anal. Chem. **75:** 1155–1163
64 Tabb D. L., McDonald W. H. and Yates J. R. III (2002) DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J. Proteome Res. **1:** 21–26
65 Han D. K., Eng J., Zhou H. and Aebersold R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat. Biotechnol. **19:** 946–951
66 Eddes J. S., Kapp E. A., Frecklington D. F., Connolly L. M., Layton M. J., Moritz R. L. et al. (2002) CHOMPER: a bioinformatic tool for rapid validation of tandem mass spectrometry search results associated with high-throughput proteomic strategies. Proteomics **2:** 1097–1103
67 Nesvizhskii A. I., Keller A., Kolker E. and Aebersold R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. **75:** 4646–4658
68 Keller A., Nesvizhskii A. I., Kolker E. and Aebersold R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. **74:** 5383–5392
69 MacCoss M. J., Wu C. C. and Yates J. R. III (2002) Probability-based validation of protein identifications using a modified SEQUEST algorithm. Anal. Chem. **74:** 5593–5599
70 Anderson D. C., Li W., Payan D. G. and Noble W. S. (2003) A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores. J. Proteome Res. **2:** 137–146

71 Fenyo D. and Beavis R. C. (2003) A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Anal. Chem. **75:** 768–774
72 Kislinger T., Rahman K., Radulovic D., Cox B., Rossant J. and Emili A. (2003) PRISM, a generic large scale proteomic investigation strategy for mammals. Mol. Cell Proteomics **2:** 96–106
73 Gygi S. P., Rist B., Gerber S. A., Turecek F., Gelb M. H. and Aebersold R. (1999) Quantitative analysis of protein mixtures using isotope coded affinity tags. Nat. Biotechnol. **17:** 994–999
74 Oda Y., Owa T., Sato T., Boucher B., Daniels S., Yamanaka H. et al. (2003) Quantitative chemical proteomics for identifying candidate drug targets. Anal. Chem. **75:** 2159–2165
75 Zhou H., Ranish J. A., Watts J. D. and Aebersold R. (2002) Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. Nat. Biotechnol. **20:** 512–515
76 Munchbach M., Quadroni M., Miotto G. and James P. (2000) Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. Anal. Chem. **72:** 4047–4057
77 Liu Y., Patricelli M. P. and Cravatt B. F. (1999) Activity-based protein profiling: the serine hydrolases. Proc. Natl. Acad. Sci. USA **96:** 14694–14699
78 Greenbaum D., Medzihradszky K. F., Burlingame A. and Bogyo M. (2000) Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. Chem. Biol. **7:** 569–581
79 Zhou H., Watts J. D. and Aebersold R. (2001) A systematic approach to the analysis of protein phosphorylation. Nat. Biotechnol. **19:** 375–378
80 Oda Y., Nagasu T. and Chait B. T. (2001) Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. Nat. Biotechnol. **19:** 379–382
81 Zhang H., Li X., Martin D. B. and Aebersold R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat. Biotechnol. **21:** 660–666
82 Ong S. E., Blagoev B., Kratchmarova I., Kristensen D. B., Steen H., Pandey A. et al. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell Proteomics **1:** 376–386
83 Oda Y., Huang K., Cross F. R., Cowburn D. and Chait B. T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. Proc. Natl. Acad. Sci. USA **96:** 6591–6596
84 Smith R. D., Pasa-Tolic L., Lipton M. S., Jensen P. K., Anderson G. A., Shen Y. et al. (2001) Rapid quantitative measurements of proteomes by Fourier transform ion cyclotron resonance mass spectrometry. Electrophoresis **22:** 1652–1668
85 Wang Y. K., Ma Z., Quinn D. F. and Fu E. W. (2002) Inverse 15N-metabolic labeling/mass spectrometry for comparative proteomics and rapid identification of protein markers/targets. Rapid Commun. Mass Spectrom. **16:** 1389–1397
86 Conrads T. P., Alving K., Veenstra T. D., Belov M. E., Anderson G. A., Anderson D. J. et al. (2001) Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and 15N-metabolic labeling. Anal. Chem. **73:** 2132–2139
87 Krijgsveld J., Ketting R. E., Mahmoudi T., Johansen J., Artal-Sanz M., Verrijzer C. P. et al. (2003) Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. Nat. Biotechnol. **21:** 927–931
88 Mirgorodskaya O. A., Kozmin Y. P., Titov M. I., Korner R., Sonksen C. P. and Roepstorff P. (2000) Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (18)O-labeled internal standards. Rapid Commun. Mass Spectrom. **14:** 1226–1232
89 Yao X., Freas A., Ramirez J., Demirev P. A. and Fenselau C. (2001) Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus. Anal. Chem. **73:** 2836–2842
90 Yao X., Afonso C. and Fenselau C. (2003) Dissection of proteolytic ^{18}O labeling: endoprotease-catalyzed 16O-to-18O exchange of truncated peptide substrates. J. Proteome Res. **2:** 147–152
91 Heller M., Mattou H., Menzel C. and Yao X. (2003) Trypsin catalyzed ^{16}O-to-^{18}O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers. J. Am. Soc. Mass Spectrom. **14:** 704–718
92 Gerber S. A., Rush J., Stemman O., Kirschner M. W. and Gygi S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad. Sci. USA **100:** 6940–6945
93 Rigaut G., Shevchenko A., Rutz B., Wilm M., Mann M. and Seraphin B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nat. Biotechnol. **17:** 1030–1032
94 Puig O., Caspary F., Rigaut G., Rutz B., Bouveret E., Bragado-Nilsson E. et al. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods **24:** 218–229
95 Mering C. von, Krause R., Snel B., Cornell M., Oliver S. G., Fields S. et al. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. Nature **417:** 399–403
96 Gavin A. C., Bosche M., Krause R., Grandi P., Marzioch M., Bauer A. et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature **415:** 141–147
97 Ho Y., Gruhler A., Heilbut A., Bader G. D., Moore L., Adams S. L. et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature **415:** 180–183
98 Blagoev B., Kratchmarova I., Ong S. E., Nielsen M., Foster L. J. and Mann M. (2003) A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. Nat. Biotechnol. **21:** 315–318
99 Ranish J. A., Yi E. C., Leslie D. M., Purvine S. O., Goodlett D. R., Eng J. et al. (2003) The study of macromolecular complexes by quantitative proteomics. Nat. Genet. **33:** 349–355
100 Neubauer G., King A., Rappsilber J., Calvio C., Watson M., Ajuh P. et al. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. Nat. Genet. **20:** 46–50
101 Neubauer G., Gottschalk A., Fabrizio P., Seraphin B., Luhrmann R. and Mann M. (1997) Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. Proc. Natl. Acad. Sci. USA **94:** 385–390
102 Rout M. P., Aitchison J. D., Suprapto A., Hjertaas K., Zhao Y. and Chait B. T. (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. J. Cell Biol. **148:** 635–651
103 Andersen J. S., Lyon C. E., Fox A. H., Leung A. K., Lam Y. W., Steen H. et al. (2002) Directed proteomic analysis of the human nucleolus. Curr. Biol. **12:** 1–11
104 Mann M. and Jensen O. N. (2003) Proteomic analysis of post-translational modifications. Nat. Biotechnol. **21:** 255–261
105 MacCoss M. J., McDonald W. H., Saraf A., Sadygov R., Clark J. M., Tasto J. J. et al. (2002) Shotgun identification of protein modifications from protein complexes and lens tissue. Proc. Natl. Acad. Sci. USA **99:** 7900–7905
106 Hansen B. T., Jones J. A., Mason D. E. and Liebler D. C. (2001) SALSA: a pattern recognition algorithm to detect elec-

trophile-adducted peptides by automated evaluation of CID spectra in LC-MS-MS analyses. Anal. Chem. **73:** 1676–1683
107 Pandey A., Podtelejnikov A. V., Blagoev B., Bustelo X. R., Mann M. and Lodish H. F. (2000) Analysis of receptor signaling pathways by mass spectrometry: identification of Vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. Proc. Natl. Acad. Sci. USA **97:** 179–184
108 Steen H., Kuster B., Fernandez M., Pandey A. and Mann M. (2002) Tyrosine phosphorylation mapping of the epidermal growth factor receptor signalling pathway. J. Biol. Chem. **277:** 1031–1039
109 Pandey A., Fernandez M., Steen H., Blagoev B., Nielsen M. M., Roche S. et al. (2000) Identification of a novel immunoreceptor tyrosine-based activation motif-containing molecule, STAM2, by mass spectrometry and its involvement in growth factor and cytokine receptor signalling pathways. J. Biol. Chem. **275:** 38633–38639
110 Goshe M. B., Veenstra T. D., Panisko E. A., Conrads T. P., Angell N. H. and Smith R. D. (2002) Phosphoprotein isotope-coded affinity tags: application to the enrichment and identification of low-abundance phosphoproteins. Anal. Chem. **74:** 607–616
111 Goshe M. B., Conrads T. P., Panisko E. A., Angell N. H., Veenstra T. D. and Smith R. D. (2001) Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses. Anal. Chem. **73:** 2578–2586
112 Ficarro S. B., McCleland M. L., Stukenberg P. T., Burke D. J., Ross M. M., Shabanowitz J. et al. (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. Nat. Biotechnol. **20:** 301–305
113 Peng J., Schwartz D., Elias J. E., Thoreen C. C., Cheng D., Marsischky G. et al. (2003) A proteomics approach to understanding protein ubiquitination. Nat. Biotechnol. **21:** 921–926
114 Kaji H., Saito H., Yamauchi Y., Shinkawa T., Taoka M., Hirabayashi J. et al. (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. Nat. Biotechnol. **21:** 667–672
115 Kjeldsen F., Haselmann K. F., Budnik B. A., Sorensen E. S. and Zubarev R. A. (2003) Complete characterization of post-translational modification sites in bovine milk protein PP3 by tandem mass spectrometry with electron capture dissociation as the last stage. Anal. Chem. **75:** 2355–2361
116 Mirgorodskaya E., Roepstorff P. and Zubarev R. A. (1999) Localization of O-glycosylation sites in peptides by electron capture dissociation in a fourier transform mass spectrometer. Anal. Chem. **71:** 4431–4436
117 Zubarev R. A., Kelleher N. L. and McLafferty F. W. (1998) Electron capture dissociation of multiply charged protein cations: a nonergodic process. J. Am. Chem. Soc. **120:** 3265– 3266
118 Sze S. K., Ge Y., Oh H. and McLafferty F. W. (2002) Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. Proc. Natl. Acad. Sci. USA **99:** 1774–1779
119 Shi S. D. H., Hemling M. E., Carr S. A., Horn D. M., Lindh I. and McLafferty F. W. (2001) Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. Anal. Chem. **73:** 19–22
120 Aebersold R. and Watts J. D. (2002) The need for national centers for proteomics. Nat. Biotechnol. **20:** 651–651
121 Mann M. (2002) A home for proteomics data? Nature **420:** 21–21
122 Aebersold R. (2003) Constellations in a cellular universe. Nature **422:** 115–116
123 Goodlett D. R., Bruce J. E., Anderson G. A., Rist B., Pasa-Tolic L., Fiehn O. et al. (2000) Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. Anal. Chem. **72:** 1112–1118
124 Smith R. D., Anderson G. A., Lipton M. S., Pasa-Tolic L., Shen Y., Conrads T. P. et al. (2002) An accurate mass tag strategy for quantitative and high-throughput proteome measurements. Proteomics **2:** 513–523
125 Lipton M. S., Pasa-Tolic L., Anderson G. A., Anderson D. J., Auberry D. L., Battista J. R. et al. (2002) Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. Proc. Natl. Acad. Sci. USA **99:** 11049–11054
126 Stritmatter E. F., Ferguson P. L., Tang K. and Smith R. D. (2003) Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. J. Am. Soc. Mass Spectrom. **14:** 980–991
127 Chaurand P., Schwartz S. A. and Caprioli R. M. (2002) Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections. Curr. Opin. Chem. Biol. **6:** 676–681
128 Chaurand P., Schwartz S. A. and Caprioli R. M. (2004) Assessing patterns in disease using imaging mass spectrometry. J. Proteome Res. **3:** 245–252