JOURNAL OF MASS SPECTROMETRY, VOL. 33, 1–19 (1998)

SPECIAL FEATURE: TUTORIAL

Mass Spectrometry and the Age of the Proteome

John R. Yates, III*
Department of Molecular Biotechnology, School of Medicine, University of Washington, Seattle, Washington 98195-7730, USA

Mass spectrometry has become an important technique to correlate proteins to their genes. This has been achieved, in part, by improvements in ionization and mass analysis techniques concurrently with large-scale DNA sequencing of whole genomes. Genome sequence information has provided a convenient and powerful resource for protein identification using data produced by matrix-assisted laser desorption/ionization time-of-flight (MALDI/TOF) and tandem mass spectrometers. Both of these approaches have been applied to the identification of electrophoretically separated protein mixtures. New methods for the direct identification of proteins in mixtures using a combination of enzymatic proteolysis, liquid chromatographic separation, tandem mass spectrometry and computer algorithms which match peptide tandem mass spectra to sequences in the database are also emerging. This tutorial review describes the principles of ionization and mass analysis for peptide and protein analysis and then focuses on current methods employing MALDI and electrospray ionization for protein identification and sequencing. Database searching approaches to identify proteins using data produced by MALDI/TOF and tandem mass spectrometry are also discussed. © **1998 John Wiley & Sons, Ltd.**

J. Mass Spectrom. Vol. **33**, 1–19 (1998)

KEYWORDS: proteome; proteins; peptides; matrix-assisted laser desorption/ionization mass spectrometry; tandem mass spectrometry

INTRODUCTION

A new paradigm has emerged for the study of living organisms. This new approach uses automated DNA sequencing to define the entire genome for an organism and it builds both an information and reagent database (Fig. 1). The genomes of 14 single-cell organisms have now been completed.[1–7] Advances in approaches for rapid sequencing of bacterial genomes have created a cottage industry for the sequencing of pathogens and single-cell organisms with interesting physiologies or unusual biological capability (e.g. *Deinococcus radiodurans* and *Methanobacterium thermoautotrophicum*).[1,8,9] Bacterial genome projects are now vigorously pursued in both the private and public sectors. Once a genome sequence has been completed, genetic and biochemical tools can be used to study quickly and, in some cases, globally the biology of the organism. The availability of complete genome sequences has dramatically altered the breadth and scale of biological experiments.

The expression of a gene results in the production of mRNA that is then translated into a protein. Processing and regulation can occur at both the transcription and translation level to control the amount of protein produced. Three key technologies have emerged as important tools to probe cells rapidly at the mRNA and protein levels. The first technology uses the defined gene sequences of an organism to construct ordered and addressable microarrays of short pieces of DNA from the organism (Fig. 1).[10,11] The expressed genes or mRNA transcripts for an entire cell are collected, converted to cDNA using the enzyme reverse transcriptase and then hybridized to the microarray. By measuring the amounts of cDNA that hybridizes to the arrayed genes, the level of expression for all the genes of the organism can be determined simultaneously. Changes in gene expression can be easily observed for each biological experiment by globally monitoring the levels of mRNA produced.

The second technology, two-dimensional gel electrophoresis, can be used to observe the expression of proteins. Large numbers of proteins can be separated, but since this is done in a collective and unaddressed fashion, the identity of any individual protein or 'spot' is not known (Fig. 1). Mass spectrometry, the third key technology, has become an important tool to correlate proteins to their genes. Mass spectrometry has had a long-standing role in the analysis of protein sequence and structure, and its role has recently expanded as an unexpected beneficiary of the information produced by whole genome analysis.[12] The definition of a gene

* Correspondence to: J. R. Yates, III, Department of Molecular Biotechnology, School of Medicine, University of Washington, Seattle, Washington 98195-7730, USA. E-mail: jyates@u.washington.edu

CCC 1076–5174/98/010001–19 $17.50

Received 4 November 1997
Accepted 20 November 1997

sequence is only part of the biological equation. The capability to identify and characterize proteins quickly is important to understanding the function of the protein encoded by the gene. Once genes are transcribed they are edited and translated into proteins. Covalent modifications to the amino acid sequence can occur co-translationally or post-translationally and neither event is dictated, at least not in a manner fully understood, by the nucleotide sequence of the gene. Some recognition sequences for sites of covalent modification (e.g. glycosylation, prenylation) in proteins are known, but the presence of a modification usually must be confirmed experimentally. Furthermore, enzyme activity may be regulated by reversible covalent modification of the protein structure. Other types of processing events, such as proteolytic cleavage, may need to occur before a protein is rendered fully functional. Proteins will also come together to form multi-protein complexes; hence identifying the interacting proteins of a biological process is an integral component of understanding the process and dissecting the functions of each of the proteins involved. The proteins undergoing interaction may change as a function of the physiological state and any given protein may be involved in numerous processes. Lastly, some proteins regulate the expression of genes by binding to regulatory sites on DNA. The dynamic and wide-ranging role of proteins has been the driving force for the development of rapid and easily utilized analytical tools for their study. This tutorial review describes first the principles of mass spectrometry as applied to the analysis of proteins and then focuses on the current state of the art for protein identification and sequencing.

IONS FROM MOLECULES

Intrinsic to the use of mass spectrometry for the analysis of peptides and proteins, and in fact all biomolecules, is the necessity to create gas-phase ions from polar or charged molecules.[13,14] Improvements in ionization technology have been the principal force driving the extension of mass range and resolution. An example was the arrival of fast atom bombardment (FAB) as a soft ionization technique for biomolecules. The ability to ionize large biomolecules by FAB quickly outstripped the capability of available mass analyzers. In response, high-field magnets were brought into use on magnetic sector instruments to measure the m/z values of the larger ions produced.[15] The mass range of quadrupole instruments was also extended from 1000–4000 u to take advantage of this new capability. The introduction of electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) led them quickly to supplant FAB as the principal methods for peptide and protein ionization.[16,17] ESI, a continuous ionization method, has been readily adapted to quadrupoles, time-of-flight (TOF), ion trap (ITMS) and Fourier transform ion cyclotron mass spectrometry (FTMS)[18–20] With somewhat more difficulty it has been used on magnetic sector mass spectrometers. MALDI is a pulsed ionization technique that uses photons to deposit energy rapidly into a matrix and analyte mixture. Pulsed ion packets are created rather than a continuous beam of ions, which limits the type of mass spectrometers easily interfaced to this technique. Mass analyzers must be capable of collecting ions for subsequent m/z separation (ITMS and FTMS) or be capable of measuring a complete mass spectrum for each ionization event (TOF).[17,21,22]

Electrospray ionization[14]

Electrospray ionization creates ions by a potential difference placed between a capillary and the inlet to the mass spectrometer. The electric field generates charged droplets in the form of a fine mist. Either through the application of a drying gas or heat, the solvent evaporates and the driplet size decreases, eventually resulting in the formation of desolvated ions.[16] A characteristic of electrospray ionization is the formation of highly charged ions without fragmentation. This process has the effect of lowering the m/z values to a range easily measured by many different types of mass analyzers. The true molecular mass of an ion can be calculated since more than one charge state is observed. Charge state and molecular mass can also be determined when the isotope forms of the molecular ion are resolved. The m/z separation is indicative of the charge state (e.g. $+2$ ions are separated by an m/z value of 0.5). While the ESI process is tolerant to low levels of buffers, salts and detergents, these substances can form adducts with the analyte causing ambiguous molecular mass determination or they can suppress the formation of analyte ions. The observation of analyte ions occurs best when samples are free of salts, detergents and buffers. A convenient method to separate contaminated substances from the analyte is to use HPLC prior to ionization of the analyte by ESI. A strength of ESI is the ease with which it can be interfaced to separation techniques.[23] The use of ESI for peptide and protein analysis is further enhanced because peptides produced by tryptic digestion (an enzyme which cleaves after lysine or arginine residues except when followed by proline) produce primarily doubly-charged ions making calculation of the molecular mass of the peptide straightforward.[24]

MALDI

MALDI emanated from years of research into the use lasers for the ionization of biomolecules. The breakthrough came with the realization that incorporation of an analyte into the crystalline structure of small UV-absorbing molecules provided a vehicle for ions to be created from polar or charged biomolecules.[25] The organic crystals must absorb at the wavelength of the laser—generally 337 nm, that of the nitrogen laser—for ionization to occur.[26] When the laser strikes the matrix crystals, the energy deposition is thought to cause rapid heating of the crystals brought about by matrix molecules emitting absorbed energy in the form of heat. Photoionization of the matrix molecules is also known to occur.[26] The rapid heating causes sublimation of the matrix crystals and expansion of the matrix and analyte

into the gas phase. Ions may be formed through gas-phase proton-transfer reactions in the expanding gas-phase plume with photoionized matrix molecules.[27] Because analyte is incorporated into a matrix during crystallization, this process may serve to sequester the analyte from contaminants such as salts and buffers, in addition to direct excitation. MALDI is reasonably tolerant of the presence of the common components of biological buffers, but improvements in the quality of mass spectra have been observed by employing a simple wash of the crystals with cold water to remove contaminants.[28] MALDI creates primarily singly precharged ions, providing a one-to-one correspondence between ions in the mass spectrum and the peptides or proteins in the mixture.[17,29]

ORGANIZING THE MASSES

Peptide and protein analysis has been a large focus of efforts in mass spectrometry over the last 10 years and many different mass analyzers have been brought into use in this area of research. In recent years great strides have been made to adapt FTMS, TOF and ITMS for use for peptide and protein analysis. While FTMS has high performance characteristics and potential, the high cost of the magnets and the complexity of operation have limited their use to industrial laboratories for very specialized purposes or to mass spectrometry research laboratories extending the applications and performance of the instrument.[30] Triple quadrupole, TOF and quadrupole ion trap mass spectrometers are three types of mass analyzers in widespread use, primarily owing to their cost and ease of use.

Quadrupole mass spectrometers

A mainstay of research and applications in mass spectrometry over the last 15 year have been the triple-quadrupole mass spectrometer.[31] Mass separation is achieved by establishing an electric field in which ions of a certain m/z value have a stable trajectory through the field. The electric fields are created by simultaneously applying a d.c. voltage and an oscillating voltage (a.c. voltage at r.f. frequencies) on four parallel metal rods, the quadrupoles. Adjacent rods have opposite d.c. polarity. Ions move with complex trajectories containing characteristic frequencies as they drift down the axis of the array of rods. By increasing the magnitude of the d.c. and r.f. voltages while maintaining the appropriate d.c. to r.f. ratio, stable trajectories are created for ions of different m/z to pass through the quadrupole array and exit to the detector. Mass resolution is dependent on the number of r.f. cycles an ion experiences in the field. However, the more cycles an ion is subjected to, the lower is the ion transmission and the greater the loss of signal at the selected m/z value.

The mass filtering effect of quadrupoles can be viewed as a separation process. By coupling quadrupole mass filters together, a powerful approach for detailed structural analysis has been created [Fig. 2(A)]. Placing a reaction region, such as a gas-phase collision cell or surface, between the two quadrupoles allows ions to be dissociated to obtain structural information.[32,33] Typically, a gas collision cell is constructed from a quadrupole mass filter operated without a d.c. voltage on the rods. Such a device functions as a high-pass filter. All ions above a set mass value are focused and transmitted through the quadrupole. In addition, enclosing the quadrupole in a cell allows the pressure to be raised to a level that permits multiple, low-energy collisions in the range 10–40 eV. Ions undergoing multiple, low-energy collisions in a short time-frame will become sufficiently activated to fragment. The principal benefit of the quadrupole collision cell is the ability to re-focus ions scattered from collisions with the neutral gases. Both mass analyzers can be operated in different modes, allowing the creation of alternate scan modes useful for identifying ions containing specific structural features.[34] For example, the first mass analyzer can be scanned over a range of m/z values while the second is set to transmit only one m/z value. This scan mode, called a precursor ion MS/MS scan, can be used to identify ions containing a specific structural feature as described by the chosen fragment ion.

Quadrupole ion trap mass spectrometers[35]

Quadrupole ion trap mass spectrometers are mass analyzers that operate by trapping ions in a three-dimensional electric field. Ions created in external sources are focused into the ion trap using electrostatic lenses.[36] An electrostatic ion gate pulses open ($-$V) and closed ($+$V) to inject positive ions into the ion trap. The ion trap is typically filled with helium to a pressure of $\sim$1 mTorr (1 Torr = 133.3 Pa). Collisions with helium dampen the kinetic energy of the ions and serve to contract trajectories quickly toward the center of the ion trap, stabilizing the motion of the injected ions. This focusing of trapped ions toward the center of the trap is achieved through the use of an oscillating potential, called the 'fundamental r.f.,' which has a frequency of roughly 10^6 Hz and is applied to the ring electrode. A mass spectrum is acquired by sequentially causing instability in the trajectories of ions and forcing them to be ejected from the ion trap volume into a detector one m/z at a time. An auxiliary frequency can be applied to the end-caps to manipulate the motion of the ions further and allow more complex experiments by selective control of the ion population in the trap. By bringing this auxiliary frequency into resonance with the frequency of motion of ions in the trap, the kinetic energy of ions of specific m/z values can be increased to the point where they are ejected from the ion trap. The principal advantage of ion trapping mass spectrometers is that the trap is first filled with ions and then m/z analysis is performed by ejecting m/z values one at a time. Unlike beam instruments, ions are not discarded prior to detection.

Multiple stages of ion dissociation (MS^n) can be performed to obtain detailed structural information from ions [Fig. 2(B)].[37,38] Ions of a specific m/z value can be isolated by a variety of methods including forward and reverse scans which sandwich the m/z value of interest

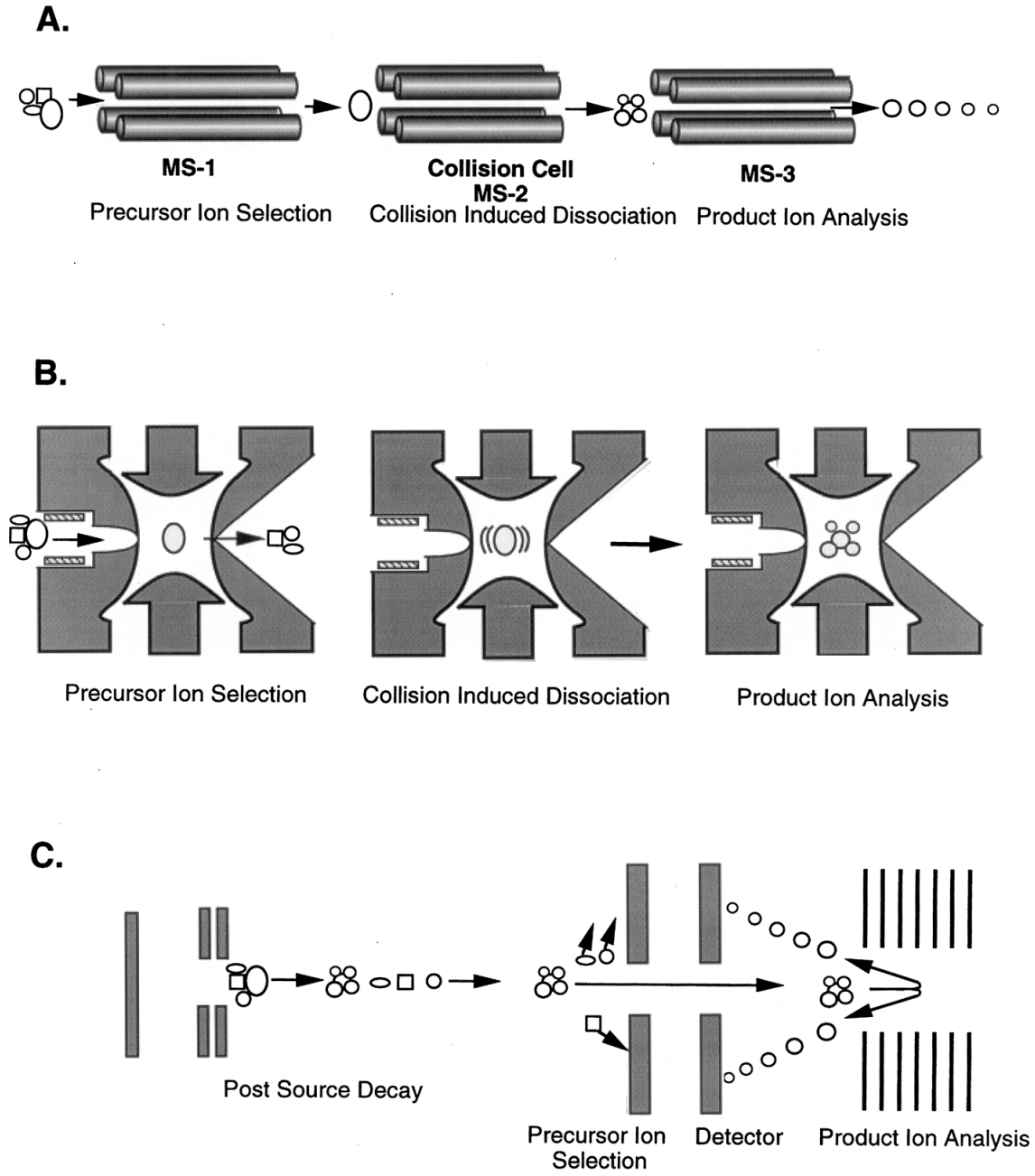

Figure 2. Three instruments used to perform tandem mass spectrometry. (A) Triple quadrupole mass spectrometer. Ions are selected with the first mass spectrometer (MS-1) and passed into a collision cell (MS-2) for activation. Fragment ions are then separated in the last mass analyzer (MS-3). (B) Quadrupole ion trap mass spectrometer. Ions are injected into the ion trap from an external ion source. A precursor ion is isolated by ejecting from the trap volume all other ions. By applying a resonance voltage across the end-caps, the motion of the selected precursor ion is increased, causing multiple low-energy collisions. The resulting fragment ions are then scanned in turn from the ion trap into a detector using a mass-selective instability scan. (C) Reflectron time-of-flight mass spectrometer. Ions are accelerated from the ion source into the field-free region. A particular m/z value is selected by timing the arrival of ions at an electronic gate. The collection of metastable decomposition products enter the reflectron. Ions are separated on the basis of kinetic energy with the lighter fragment ions exiting earlier.

between the limiting values set by two scans or by using broadband notch techniques.[38–40] The kinetic energy of the selected ion population is then increased by applying a voltage resonant with the frequency of the precursor ion, causing more energetic collisions with the He bath gas. By subjecting the ions to many hundreds of low-energy collisions, the internal energy of the ion is increased until fragmentation occurs. If a single or a very narrow resonance frequency is used for excitation, the fragment ions produced are no longer excited as their velocities will have changed. Broadband excitation methods will subject the fragment ions to additional energetic collisions.[40,41] A dissociation product can also be trapped within the ion trap volume for dissociation. An MS^3 experiment can be used to create structurally important fragment ions of a particular dissociation product from the ions produced in an MS^2 experiment. In contrast to single-frequency excitation methods, fragment ions produced in the collision cell of a triple-quadrupole mass spectrometer continue to be subjected to collisions until they exit the cell. Consequently, fragmentation patterns for peptides caused by single-

frequency excitation in the ITMS instrument can be slightly different than those observed on a triple-quadrupole mass spectrometer.

Time-of-flight mass spectrometers[42]

Time-of-flight mass spectrometers are among the simplest of the mass analyzers. Mass-to-charge ratios are determined by measuring the time it takes for ions to move through a field-free region. Given a constant accelerating voltage, the flight time for an ion is related to its m/z ratio. The flight path for an ion can be increased, without unduly increasing the size of the flight tube, by incorporating an ion mirror or reflectron at the end of the flight tube. Ion direction is reversed to send the ions back down the same vacuum chamber at a slightly different angle so the flight path of the reflected ions does not cross with the ions entering the reflectron. Most importantly, a reflectron can also correct for minor kinetic energy differences among ions of the same m/z value and so minimize variations in flight times. More energetic ions arrive earlier at the reflectron and penetrate deeper before reflection than the less energetic ions. Ions of the same m/z but different initial energies then meet at the detector. In MALDI, one contribution to the kinetic energy distribution of ions is thought to result from their acceleration through the gas-phase plume created during desorption.[43] Brown and Lennon[44] have observed that both mass resolution and mass measurement accuracy can be improved by delaying the extraction of ions from the ion source allowing correlation of the space and velocity components of the desorbed ions. The ion packet is better focused as it is accelerated from the source resulting in much improved performance. MALDI/TOF mass spectrometers have the ion source in-line with the flight tube, a design well suited to the pulsed nature of the ionization technique. To interface ESI to TOF mass spectrometers, the ion source is placed perpendicular to the flight tube.[18] As the ions exit the source they are pulsed into the flight tube and mass analyzed. This design has been particularly useful in coupling electrospray ionization to TOF.[18,45,46]

Mass measurement of peptides and proteins has been improved by advances in TOF technology and instruments. Structural information, however, is deduced from ion fragmentation. Two different types of fragmentation events have been observed in TOF mass spectrometers using MALDI. The first is metastable decomposition occurring in the field-free region of the TOF flight tube. This process is called post-source decay (PSD) [Fig. 2(C)].[47,48] Fragments created in the field-free region of the flight tube have the same velocity (but kinetic energies which vary with their masses) as the precursor ion, since the fragmentation event occurs outside the accelerating field of the ion source. Fragment ions of higher m/z value will penetrate deeper into the reflectron and exit later than the lighter product ions. A separation based on m/z value is thereby effected in the reflectron. The second process, observed by Lennon and Brown,[49] involves fast metastable decay in the ion source and can be observed when a long delay (320 ns) is instituted before ion acceleration. Fragment ions are produced prior to acceleration from the ion source and consequently the fragments have the normal high kinetic energy as they leave the source and undergo separation in the field-free region of the mass spectrometer. This fragmentation process is inefficient and forms Cn-, Yn- and Z*n-type ion fragments, but sufficient sequence information can be acquired for partially sequencing a protein. This process has been observed for ions of small proteins, but it is still uncertain how amino acid sequence and protein size affect it. This method of dissociating proteins, however, has exciting potential. A MALDI/TOF mass spectrometer can also be fitted with a collision cell to induce peptide ions to fragment. In-source collision-induced dissociation (CID) can be performed on ESI/TOF mass spectrometers by employing higher potential differences in the source to induce energetic collisions. By inducing fragmentation of peptide ions prior to entering the flight tube, however, the potential for mixture analysis is decreased because individual components are dissociated prior to an ion selection stage.

DELIVERING PEPTIDES TO THE MASS SPECTROMETER

An important aspect of the use of mass spectrometry for biological studies is the method used to introduce samples into the mass spectrometer. Moving and manipulating small quantities of protein from the laboratory bench to the analytical instrument requires care and thought to minimize sample losses en route. Several strategies have been developed based primarily on the method of sample introduction into the ionization source. In general, biological experiments produce proteins or peptides for analysis in relatively dilute form and with a background of buffers and detergents. Both MALDI and ESI require approaches tailored to the particular ionization technique. ESI is a liquid introduction technique, and consequently samples can be manipulated exclusively in the liquid state. Analysis by MALDI, however, requires that samples be deposited on a sample plate and then co-crystallized with a matrix. In both techniques the concentration of the sample and the complexity of the background matrix influence the sensitivity that can be achieved.

Two general approaches are used in MALDI to prepare samples in optimum form for analysis. Typically 1–2 μl of solution are deposited in the sample plate. For very dilute solutions, the sample needs to be concentrated. Traditional sample concentration techniques such as lyophilization can be used, but the risk is run that small quantities of sample will be lost to the walls of the container. In addition, the contaminants will be concentrated along with the sample. Except in rare cases, small quantities of sample should never be dried. Reversed-phase high-performance liquid chromatography (HPLC) can be used to remove buffers and salts, and to separate the analyte from contaminant. Complex peptide mixtures can be separated by HPLC and fractions collected for analysis. The peptides are often collected in fairly large volumes relative to the 1–2 μl required for sample deposition and this approach

does not take advantage of the strength of MALDI/TOF for the analysis of unfractionated peptide mixtures. A variation of this approach involves the absorption of peptides to a small quantity of reversed-phase packing material and then batch elution in a small volume of liquid, generally corresponding to the volume of solution required for deposition on the sample plate. This approach can serve to concentrate and clean the bulk sample and to focus the sample to a small area. Alternatively, the plate can be coated with a thin layer of nitrocellulose to bind samples before washing away contaminants prior to depositing matrix.[50] The cross-section of the laser beam will cover an area of $\sim$50–100 μm^2 and the signal-to-noise ratio can be optimized by minimizing the ratio of sample surface area to the area impinged on by the laser. Sample manipulation and deposition methods are decoupled from the actual sample analysis step, so samples can be prepared prior to the analysis. The process is also amenable to automation using robotics to prepare and deposit the samples.[51]

ESI enables samples to be manipulated on-line in conjunction with many separation or solution delivery methods. The most notable are the separation techniques of HPLC and capillary electrophoresis (CE). Both of these techniques simultaneously combine analyte preconcentration and purification. At low flow rates (<100 $\mu l\ min^{-1}$), ESI behaves as a concentration-dependent device, thus minimization of the column elution volume maximizes the concentration of the peptide subjected to ESI. Low flow-rate separation techniques such as microcolumn HPLC and CE are ideal for integration with ESI to maximize sensitivity of analysis. Kennedy and Jorgenson[52] and Flurer *et al.*[53] have developed methods to create microcolumns on the 50–100 μm scale. To use μm scale microcolumns for separations requires that reversed-phase solvent gradients be formed at high flow-rates, followed by flow splitting to reduce the flow-rate to sub-$\mu l\ min^{-1}$ levels. An alternative to flow splitting is to pre-form and store a solvent gradient and then deliver the gradient by pushing the solvent through the storage loop and on to the column.[54] Integration of microelectrospray ionization or very low flow-rate electrospray ionization has further increased the sensitivity of microcolumn HPLC.[55] By using electrospray emitters, columns with very small diameter dips, stop-flow techniques can be implemented to increase the length of time ions can be electrosprayed.[55,56] When the solvent flow and gradient are stopped, the sample contained in the tip of the emitter continues to flow because the tip of the emitter is the site of greatest pressure restriction. This procedure gives added time to acquire tandem mass spectra for co-eluting ions or to signal average for a greater length of time to improve the signal-to-noise ratio. A recent comparison of the 'peak parking' or stop-flow analysis method to nanospray infusion showed improved performance for the acquisition of mass spectral data from complex mixtures of peptides.[56]

Capillary electrophoresis is another technique well suited for the separation of peptides prior to electrospray ionization. Separations in CE are driven by electroosmotic flow created by the surface dielectric present when a high potential (10–30 kV) is dropped across a fused-silica column.[57] Peptides have different migration profiles based on their electrophoretic mobility and thus can be resolved over the course of the separation. In addition, the flow profile of peaks in CE are flat rather than Gaussian and this results in very narrow bands that deliver high concentrations of sample to the mass spectrometer and so improve the sensitivity in electrospray ionization. There is a trade-off as peaks that are too narrow can be missed if examined at too slow a mass scan speed. Smith and co-workers[58] observed that smaller column diameters result in lower flow-rates and consequently can achieve attomole levels of sensitivity when interfaced to a mass spectrometer. To maximize the formation of peptide ions by ESI, separations are best conducted at low pH. By derivatizing the walls of the fused-silica column to reverse the surface charge (negative to positive), peptide interactions with the surface are decreased and electroosmotic flow is increased at low pH.[59,60] CE requires the injection of small volumes (pico- to nanoliters) of sample to maintain peak shape and resolution and this requirement forces the use of highly concentrated samples. Two approaches to sample preconcentration have been developed to overcome this limitation. The first method concentrates dilute samples on a small column (1–2 cm) of reversed-phase HPLC packing material placed near the head of the CE column.[61–63] The second method is similar in principle but uses a hydrophobic membrane inserted into the flow path of the CE column near the head of the column.[64,65] After the analytes have been concentrated on the solid support, an aliquot of organic solvent is used to elute them in a small volume to perform the CE separation. Very dilute samples can be concentrated at the head of the column. Reduction of these techniques to practice has greatly improved the effectiveness of CE in combination with tandem mass spectrometry for peptide analysis.[66]

A third strategy for introduction of peptides uses low flow-rate infusion coupled to off-line or on-line concentration and batch elution. This strategy attempts to infuse a collection of peptides rather than separate them by HPLC or CE. Gale *et al.*[67] reported that low flow-rate ESI (micro-electrospray source) could be used to spray small volumes of aqueous solutions efficiently.[67] This ion source could achieve pressure-assisted flow rates as low as 200 $nl\ min^{-1}$ with etched fused-silica capillaries with inner diameters as small as 20 mm. They demonstrated improvements in electrospray signal stability and a 4–10-fold improvement in sensitivity, in addition to a 2–5-fold decrease in sample consumption. Another method, used by Emmett and Caprioli,[68] achieved similar improvements in sensitivity, signal stability and signal-to-noise ratio by combining on-line clean-up of the sample through a reversed-phase packing material and then batch elution of the peptides. This micro-electrospray source was used to achieve detection limits of 1 fmol (10^{-15} mol) (total sample consumed) for small peptides such as methionine-enkephalin. Wilm *et al.*[69] used the same principle of low flow-rate infusion, but employed off-line sample clean-up through reversed-phase packing material. By pulling borosilicate glass capillaries to a very fine tip, the flow-rate was reduced to a low $nl\ min^{-1}$ rate.[70] These sample introduction and ion source designs can also be

beneficial when electrospraying small quantities of protein and peptide mixtures in small volumes (typically 1–2 μl of solution) for extended periods of time (several tens of minutes) to perform a variety of mass spectrometric experiments. In addition, this extended analysis time helps in achieving acceptable signal-to-noise ratios by performing extensive signal averaging on peptides present in small quantities. A drawback to this approach is the presence of contaminants or a high background that impedes the effective identification of peptide ions for MS/MS analysis. To identify relevant peptide ions, precursor ion MS/MS scans can be employed to identify ions containing features unique to peptides such as immonium ions for common amino acid residues, e.g. Leu and Ile.[71]

MASS ANALYSIS OF PEPTIDES AND PROTEINS

Protein analysis can be divided into two basic categories: mass analysis and amino acid sequencing. An increased emphasis on mass measurement began with the development of FAB ionization. This trend coincided with breakthroughs in molecular biology and DNA sequencing that created a shift in approaches for protein sequencing. Early DNA sequencing methods were error prone and, therefore, a need developed to check quickly protein sequences translated from nucleotide sequences.[72,73] The ability to determine rapidly m/z values for the collection of peptides obtained by proteolytic digestion of a protein provided a fast method to check the fidelity of a translated sequence or to determine the reading frame of the gene (Fig. 3). If the protein was purified from native cell extracts, the locations of post-translational modifications could also be determined. Mass measurement alone could provide sufficient information to validate a translated sequence. The introduction of MALDI and ESI techniques permitted significant changes in strategies for analysis of proteins in addition to creating the ability to measure accurately the mass of the intact protein. The ability to determine accurately the mass of intact proteins provided a method to check quickly if a protein sequence was correct or to determine the presence of post-translational modifications. If the measured mass deviates from the expected mass, then either the sequence is incorrect or the protein contains modifications. Obviously, an initial mass measurement for an unknown protein would become more useful once the protein had been identified or sequenced. Efforts have been made to measure the molecular masses of proteins separated by gel electrophoresis directly by using MALDI/TOF and they are meeting with increasing success. In this method, the proteins are electroblotted to a suitable membrane or cut directly from the polyacrylamide gel slice.[74–77] The dehydrated gel slice or the membrane is placed on the sample stage, treated with a matrix solution and irradiated with the laser. The protein desorbs from the gel or membrane and the molecular mass is determined. To identify the protein after a molecular mass has been measured, an *in situ* proteolytic digestion can be performed and the m/z values of the resulting peptides measured or tandem mass spectra acquired.[78] This approach, although still rather cumbersome, has the advantage of combining the separation power of gel electrophoresis with the mass measurement accuracy of mass spectrometry.

Sequencing peptides by mass analysis of sequence ladders

Sequence information is not *a priori* determined from a mass measurement for either a peptide or a protein. Consequently, chemical and enzymatic strategies have been coupled with mass measurement to create sequence ladders from sequentially and incompletely shortened peptides. Two approaches have appeared. The first approach emerged in the early 1980s using FAB to create ions from peptides sequentially shortened through the use of carboxypeptidase Y to leave a ragged *C*-terminus.[79] This approach has been resurrected and used in conjunction with highly sensitive MALDI/TOF.[80,81] By adjusting the concentration of the enzyme used to create the ladder, the rapid enzymatic cleavage process can be slowed and ions arising from cleavage of each amide bond can be observed. The process of ladder sequencing depends on an incomplete reaction and the enzymatic reactions can be hard to control or they may be sensitive to the *C*-terminal sequence of the peptide. A second approach uses an Edman degradation chemical cleavage strategy with a small percentage of a chemical terminator added to the reagents.[82] A small amount of the peptide's *N*-terminus is blocked at each coupling step to prevent further degradation. After a sufficient number of cycles have been performed, an aliquot is removed for mass analysis. By subtracting the difference of adjacent m/z values, sequence information is obtained. The chemical background can be high as unreacted chemicals and side products build up from the repeated cycles although volatile reagents for ladder sequencing have been developed to circumvent this problem.[83] Ladder sequencing strategies, however, suffer from the requirement of a pure or nearly pure peptide to be successful, but this is a problem that can be partially overcome with improved mass accuracy and resolution in the mass measurement process.

Correlating peptide m/z information data with known sequences

The intent of protein analysis strategies is to derive sufficient information to identify the protein. Over the last 10 years, this has meant obtaining sufficient sequence information to clone the gene. In recent years, the first step after obtaining sequence is to search the databases to determine if the sequence is known. In 1993 five laboratories independently developed computer algorithms to use peptide mass maps to search protein databases and so identify proteins.[84–88] The observed m/z values are compared with the values predicted after digestion of each protein in the database with a site specific enzyme (Fig. 3). Provided a sufficient number of peptide ions are observed in the mass analysis step, and

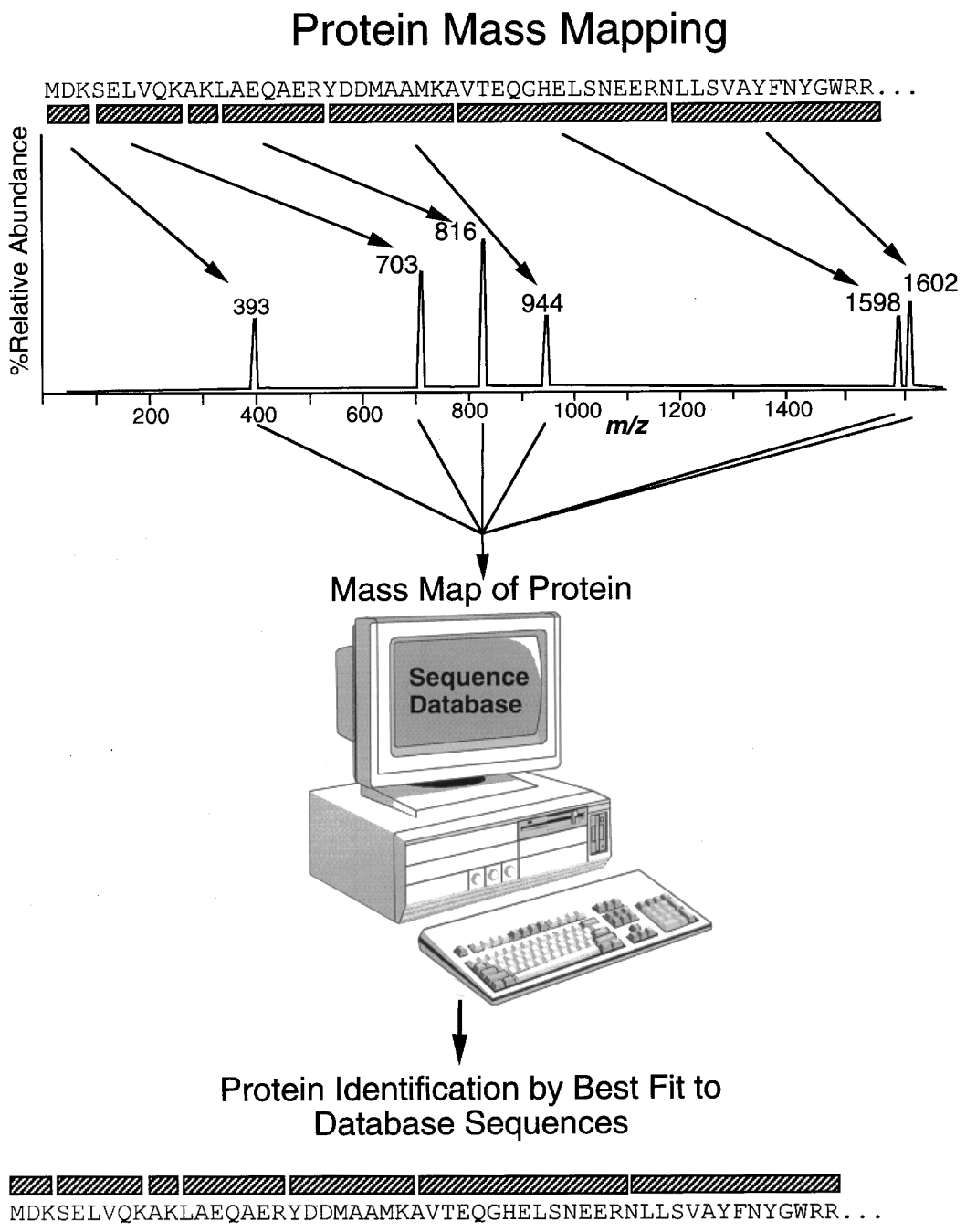

Figure 3. Peptide mass mapping. A protein sequence can be verified by site-specific digestion and measurement of the peptide ions for correlation with those predicted by the sequence. Conversely, if the identity of the protein is not known the peptide mass map can be used to search the protein database to find the sequence that best fits the mass map.

the protein is not heavily modified, and there are not more than two proteins present, a match can generally be found. This method has found significant application in the rapid identification of proteins from gels, in particular, in the identification of proteins from two-dimensional gels.[84] A level of uncertainty in the identification can be observed when searching large databases with this technique, consequently it has been combined with other types of information to increase the specificity of the identifications. James *et al.*[89] have added information from proteolytic digests of a different specificity to increase search accuracy. Clauser *et al.*[90] employed tandem mass spectrometry to sequence a peptide when an identification based on a mass map was uncertain. Lastly, sequence information, if available, can be added to the measured m/z values to increase the certainty of an identification.[91] The recent improvements in mass accuracy (10–50 ppm) and resolution (10 000–15 000) in TOF mass spectrometers created by the combination of delayed extraction and reflectrons have improved the accuracy of mass measurement, which minimizes ambiguity in the identification.[92,93] Mass mapping has also been proposed as a method for cross-species identification. This process involves using the database information of one organism to identify the similar or homologous protein

of another organism for which there may not be much protein sequence.[94,95] This approach relies on sequence conservation between two proteolytic cleavage sites (a stretch of ~10–15 amino acids) in the same protein of the two different organisms.

IDENTIFYING AND SEQUENCING PEPTIDES USING TANDEM MASS SPECTROMETRY

Methods for protein sequence analysis using tandem mass spectrometry began in the late-1970s and were improved as ionization methods more suitable for peptides became available.[96] The strategy can be employed on several different types of mass spectrometers capable of selecting single m/z values and subjecting the ions to CID. The predictable fragmentation patterns of peptides have led to the development of sequencing and identification methods using tandem mass spectra of peptides.

Peptide ions fragment primarily at amide bonds along the backbone under low-energy gas phase collision conditions, generating a ladder of sequence ions.[96,97] If the charge is retained on the *N*-terminal portion of the fragment ion after cleavage of the amide bond, then b-type ions are formed; however, if the charge is retained on the *C*-terminal portion, y-type ions are formed.[98,99] A complete series of either one or both ion types allows the determination of the amino acid sequence by subtraction of the masses of adjacent sequence ions (Fig. 4). Mechanistic studies have shown that the position of basic amino acid residues such as arginine and lysine along the peptide backbone dictates the preponderance of a single type of sequence ion. If arginine or lysine is present at the *N*-terminus, the tandem mass spectrum is dominated by b-type ions, and if these basic amino acid residues are present at the *C*-terminus, the tandem mass spectrum is dominated by y-type ions.[98,100] However, if these basic amino acid residues are somewhere in the middle of the peptide backbone, both types are common and the complexity of the spectrum is increased.

Correlating MS/MS data to sequences in databases

The rapid increase in the number of protein sequences in the database has greatly aided the analysis of tandem mass spectra of peptides. By using computer algorithms, the information created by the CID of peptides can be used to search protein and nucleotide databases.[101] Yates and co-workers[12,101–104] developed an approach to use the fragmentation information contained in a tandem mass spectrum to search protein and nucleotide databases to identify the amino acid sequence represented in the spectrum and thereby identify the protein (Fig. 4). The fragment ions contained in the spectrum provide an added degree of specificity to the peptide mass and can allow identification of proteins based on a single tandem mass spectrum. Most computer algorithms use each tandem mass spectrum in an independent search of a database. The probability that two or more tandem mass spectra, of reasonable quality, e.g. good signal-to-noise ratio, will incorrectly match to the same protein sequence is small. The independent and highly specific nature of the information contained in a tandem mass spectrum allows specific proteins to be identified among individual proteins present in mixtures, including protein contaminants such as keratin, antibodies and protease autolysis products.[102]

Tandem mass spectra of peptides contain three levels of information. First is the mass of the peptide. An accurate peptide mass alone can reduce the number of sequence possibilities to a small number (tens to thousands) when coupled with the specificity of the enzyme used to create the peptide. Without a knowledge of the specificity of the enzyme and with a wide mass tolerance, larger numbers of peptides can be identified simply on the basis of mass in large databases. For example, Yates *et al.*[105] observed a peptide of mass 1480.7 u ($\pm$3 u) will match $\sim 6.5 \times 10^6$ different amino acid sequences in a nucleotide database search when employing a six-frame translation of the nucleotide sequences to protein sequences. To identify uniquely the amino acid sequence represented by a measured mass, a second level of information is needed and a tandem mass spectrum provides a pattern of fragment or sequence ions that is fairly unique to a given sequence. By correlating the predicted fragment ions for each of the amino acid sequences matching the mass of the peptide to those observed in the tandem mass spectrum, the closeness of fit can be determined. The same group[101,102] devised a method to evaluate amino acid sequences pulled from the database by reconstruction of a model tandem mass spectrum and its comparison with the observed tandem mass spectrum using a cross-correlation function (SEQUEST). This method determines the similarity of the reconstructed spectrum to the spectrum obtained experimentally. The magnitude of the cross-correlation value indicates the quality of the match between the sequence and the spectrum and the difference between the normalized cross-correlation score to the second ranked sequence shows the quality of the match versus all the other top ranking sequences in the database. Each of the fragment ions represents part of the amino acid sequence, and consequently the pattern as a whole forms a unique signature in much the same way that a mass map can be a unique fingerprint for a protein sequence. The third level of information present in a tandem mass spectrum is the actual sequence. By interpreting a short stretch of amino acid sequence (3–4 amino acid residues) that may not be unique for a specific protein, and combining that information with the mass of the overall peptide or the masses of fragment ions, a higher level of specificity is created.[106] A fragment ion mass represents a shorter segment of the peptide's amino acid sequence and, in combination with the partial sequence, it significantly constrains the number of amino acid sequences in the database that will fit all the information. Methods for searching databases have been devised for tandem mass spectrometric data obtained from several different types of mass spectrometers.[107–111] A potential caveat when using tandem mass spectrometric data to search large databases is the existence of sequence conservation in

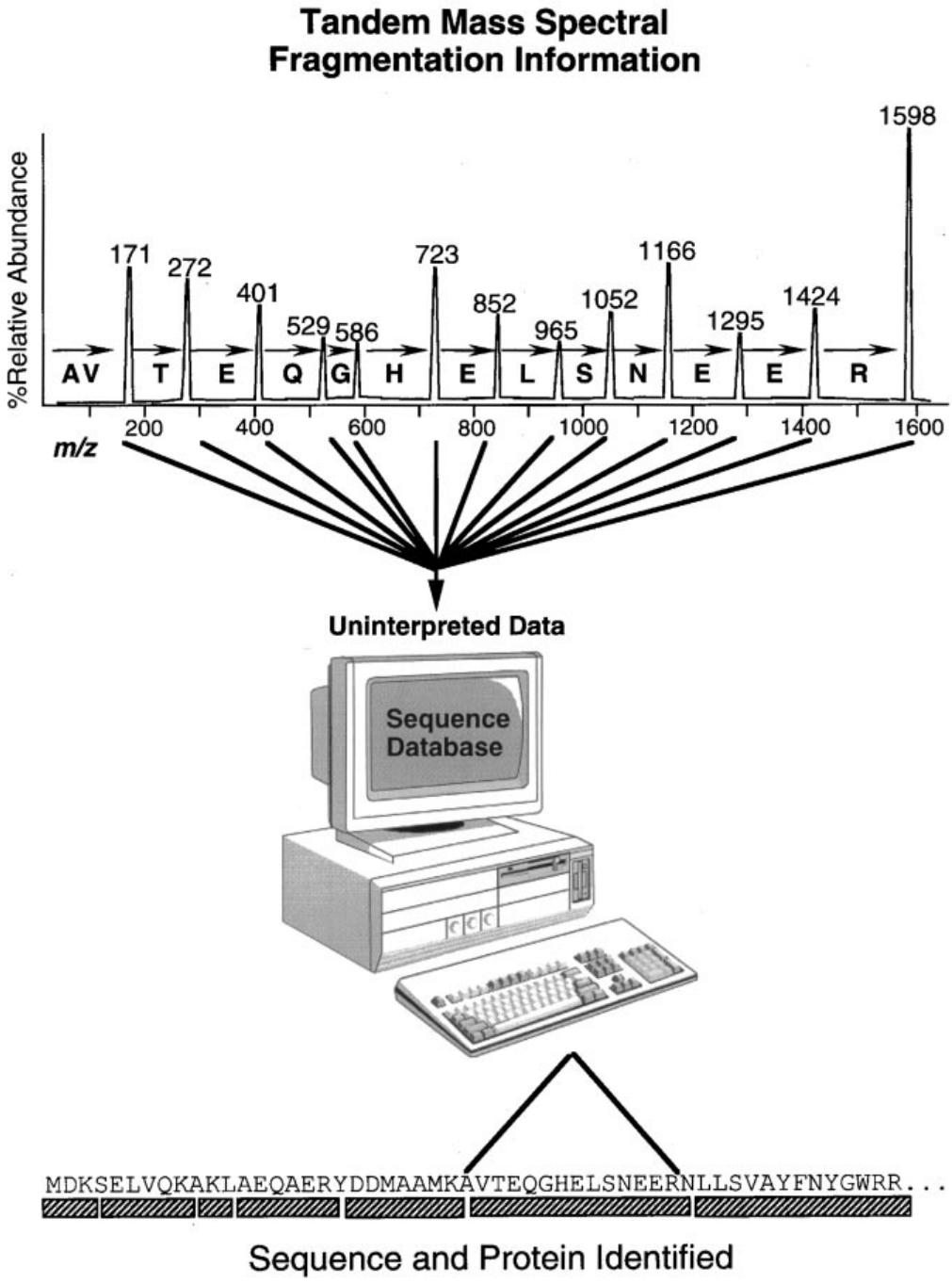

Figure 4. Tandem mass spectrometric sequencing. The ladder of fragment ions represents the amino acid sequence of the peptide. By subtracting the *m/z* values for adjacent ions of the same type the sequence can be elucidated. Conversely, the fragmentation pattern can be used to search the protein or nucleotide database to find the amino acid sequence that best fits the tandem mass spectrum.

similar or related proteins. A single tandem mass spectrum matching the same sequence in more than one protein will not provide a unique identification and additional tandem mass spectra should be used to try to differentiate among those identified proteins.

De novo sequencing of peptides

Even though the databases are filling with sequences, the need to interpret tandem mass spectra to derive a sequence still exists. CID produces a ladder of sequence ions where the difference between consecutive ions indicates the mass of the amino acid at that position in the sequence. Successful interpretation involves determining which ions originate from the *N*- or *C*-terminus so mass differences between consecutive ions of the same type can be calculated [Fig. 5(A)]. Hence a set of sequence ions, from low mass to high mass, will define the amino acid sequence.

In Fig. 5(B), a high-resolution scan of the *m/z* value of the doubly protonated precursor ion [Glu]^1-fibrinopeptide B, a common standard used for tuning

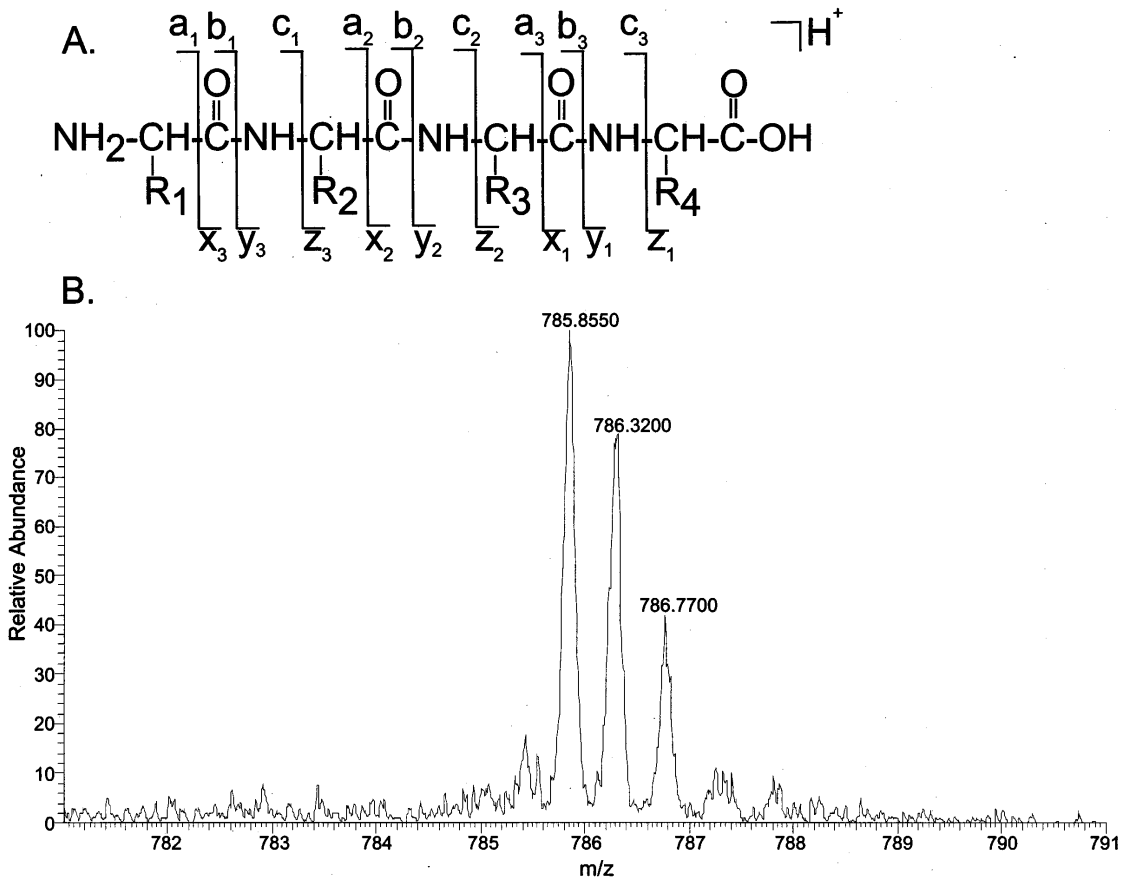

Figure 5. (A) Nomenclature for fragmentation of peptides. (B) High-resolution mass scan of the doubly charged ion of [Glu]1-fibrinopeptide B.

tandem mass spectrometers, is shown. If the molecular mass is calculated from the observed m/z (observed $M_r = 1569.716$) and compared with the predicted molecular mass of the peptide (calculated $M_r = 1569.669$), it is found to be within 0.047 u. The tandem mass spectrum acquired on a Finnigan MAT LCQ ion trap mass spectrometer is shown in Fig. 6. MS/MS analysis of peptides derived from tryptic digests of proteins generally present a prominent y-type ion series in the high-mass end of the spectrum and a Lys or Arg residue as the *C*-terminal amino acid. These amino acids can be recognized by y_1-type ions at m/z 147 or 175, but ion trap spectra of peptides do not show the low m/z range when large peptide ions undergo MS/MS. The interpretation strategy for ion trap MS/MS data is thus modified based on the data available.

The interpretation can be started with the most abundant high-m/z ion, in this case m/z 1285. A window of m/z values to search for the next sequence ion can be created by subtracting 57 (Gly) and 186 (Trp) from the ion at m/z 1285 (m/z 1009–1228). Each ion present in the window is subtracted starting with the most abundant ions present to find differences corresponding to the mass of an amino acid residue. In this case the ion at m/z 1171 shows a difference of 114.04, which corresponds to the amino acid Asn. If this process is continued through the tandem mass spectrum, a stretch of eight amino acids can be readily determined. Traditionally, amino acids with 1 u differences in their molecular masses (Asn, Asp, Leu/Ile and Glu, Gln) are difficult to distinguish. In this example, five of the eight amino acids have 1 u differences and are clearly resolved. How is the sequence completed? Knowing the peptide is produced by trypsin cleavage can narrow the *C*-terminal residue to Lys or Arg. The ion present at 246.1 is the y_2 ion, so likely amino acid combinations are Val–Lys and Ala–Arg. A check of the spectrum shows a very low abundance ion at 1396. The difference between the $[M + H]^+$ and this ion is 175, the residue mass of Arg strongly suggesting that the *C*-terminal sequence of the peptide is Ala–Arg.

To resolve the identity of the *N*-terminal amino acids is more difficult and ultimately would require an MS^3 experiment to confirm any interpretation. To gain more information about the sequence, the peptide was derivatized with methanolic HCl to convert the acidic residues into methyl esters. A mass shift of 70 u is observed for the peptide, indicating the presence of five acidic groups. One acidic residue is unaccounted for in the current sequence, consequently an Asp or Glu residue must exist in the remaining sequence. The remaining mass unaccounted by the sequence is 285 u. Subtracting the molecular masses of Asp and Glu leaves a remainder of 170 and 156.

A molecular mass of 170 could be Val–Ala or Leu/Ile–Gly and the 156 could be Val–Gly or Arg. An MS^3 experiment is needed to resolve the differences clearly,

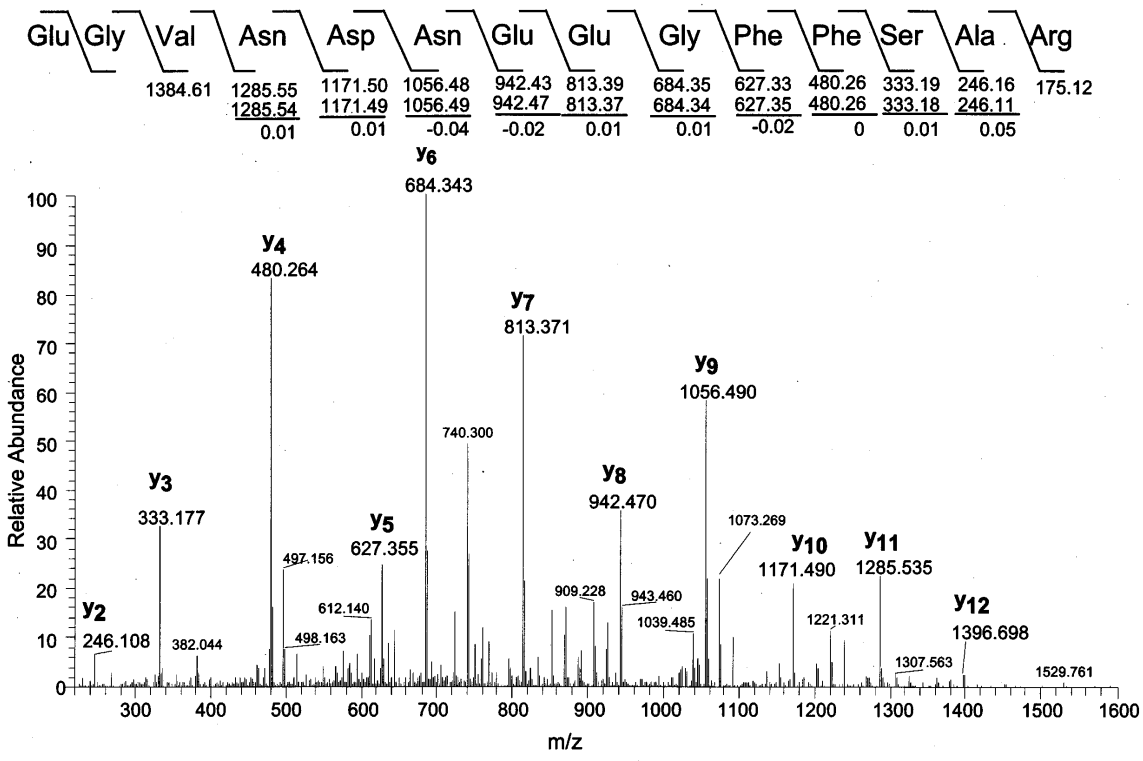

Figure 6. Collision-induced dissociation mass spectrum recorded on the $[M+2H]^{2+}$ ions at *m/z* 786 of [Glu]1-fibrinopeptide B. Fragments of type y-ions having the general formulae $H(NHCHRCO)_n^+$ and $H_2(NHCHRCO)_nOH^+$, respectively, are shown below the amino acid sequence at the top of the figure. The *m/z* values just below the sequence are those predicted from the sequence. Below the predicted values are the *m/z* values measured and the difference between the two values.

but some good possibilities can be examined. No matter what instrument is being used for tandem mass spectrometry, the possibility of acquiring incomplete sequence information is present. Based on the sequence data already available, the information can be used to search the database using sequence similarity programs or, if additional sequence has been obtained from the protein, the sequences can be used to clone the gene.

De novo sequence analysis using tandem mass spectrometry has been successfully applied to many peptides and proteins. The first protein sequenced using tandem mass spectrometry was the delta hemolysin protein.[112] Initially, tandem mass spectrometry was used alone or in conjunction with Edman degradation to sequence whole proteins or parts of proteins.[113,114] Some recent applications have focused on femtomole-level sequencing of bioactive peptides bound to MHC class I and class II molecules and of tumor antigens.[115–117] Tandem mass spectrometric sequencing, at the picomole level, to generate information for cloning proteins has also become more common.[69,118–120]

IDENTIFICATION OF ELECTROPHORETICALLY SEPARATED PROTEINS

Polyacrylamide gel electrophoresis is one of the most commonly used techniques to separate proteins. Its great strength lies in the generality of the method. A highly resolving version of gel electrophoresis is two-dimensional gel electrophoresis (2-DGE), which combines a separation by isoelectric point in the first dimension with separation by size in the second dimension. This method is capable of separating thousands of proteins in a single analysis. Traditionally, proteins separated by gel electrophoresis were identified by reaction with antibodies to proteins suspected of being present or, as a more general method, by using Edman degradation to obtain the *N*-terminal amino acid sequence. Amino acid sequencing became more effective for the identification of proteins with the development of membranes compatible with electroblotting and sequencing chemicals.[121,122] Alternative approaches for acquiring amino acid sequences, especially when the *N*-terminus of the protein is blocked, required *in situ* digestion of an electroblotted protein or digestion directly in the polyacrylamide matrix.[123] The collection of peptides produced was then separated and sequenced individually to obtain the internal sequence. Several limitations to the use of Edman sequencing for protein identification are evident. The sensitivity of Edman sequencing has improved only slowly over the last 10 years, the method is limited to proteins without blocked *N*-termini and peptides need to be purified before sequencing. To improve the identification of proteins separated by gel electrophoresis, more recent approaches have applied mass spectrometry with great impact.

Two different mass spectrometric strategies have been employed to identify proteins separated by gel electrophoresis (Fig. 7). All of the published approaches use *in*

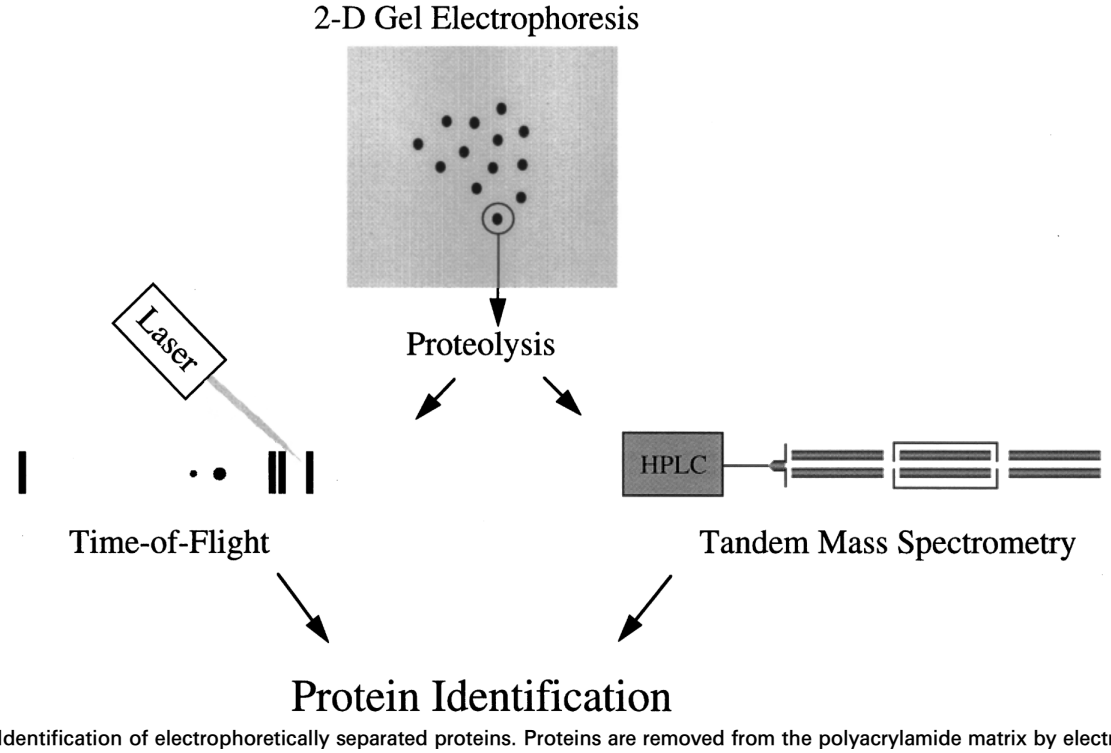

Figure 7. Identification of electrophoretically separated proteins. Proteins are removed from the polyacrylamide matrix by electroblotting to a membrane or by *in situ* digestion in the gel. The peptides are collected and analyzed by tandem mass spectrometry or by MALDI/TOF. Peptide mixtures can be introduced into the tandem mass spectrometer through HPLC, capillary electrophoresis or nanoelectrospray ionization to obtain tandem mass spectra. A MALDI/TOF mass spectrometer can be used to measure the *m/z* values for peptides. Data from the tandem mass spectrometer or MALDI/TOF can be used to search sequence databases to identify the protein.

situ digestion methods after gel electrophoresis to obtain peptides from the protein for analysis.[66,124,125] The collections of peptides obtained are suitable for mass spectrometric analysis to obtain peptide mass maps and, thus, to identify a protein after a database search. The potential for high-throughput analysis has led to a few large-scale projects, such as the identification of proteins from myocardial cells of the human heart and proteins from *Saccharomyces cerevisiae*.[126,127] When peptide mass mapping is used as the principal identification method, ambiguous identifications are confirmed or completed with the use of tandem mass spectrometry. Tandem mass spectrometry, coupled to CE or HPLC, has also been used as the sole method to identify proteins separated by 2-DGE, with excellent results.[66,128] One study of the *H. influenzae* proteome used 2-DGE to separate the proteins and tandem mass spectrometry to identify 260 proteins. Of the 260 spots analyzed and identified, 26 contained two or more proteins. Three proteins were found whose genes contained frameshift errors. These proteins were identified by using a six-frame translation of the nucleotide sequence to match tandem mass spectra to the protein. For one spot taken off the gel, no match in the *H. influenzae* database could be found for the tandem mass spectra. Upon searching the *E. coli* database, a perfect fit was found for all the peptide tandem mass spectra obtained identifying the protein as tryptophanase. No trace of the gene could be found in the genome of the Rd strain used for DNA sequencing. The NCTC 8143 strain used in the proteome analysis, however, still contained the gene. While large-scale protein identification of total cell lysates can provide insight into the accuracy of open reading frame predictions and the identification of the most highly expressed proteins, they have yet to yield new insights into key biological processes. A more informative approach is to combine subtractive analysis techniques in conjunction with 2-DGE and mass spectrometry to compare different cellular states or to use this approach to separate and analyze enriched protein fractions.

'SHOTGUN' IDENTIFICATION OF PROTEINS IN MIXTURES

Tandem mass spectrometry has the capability to analyze the individual components of mixtures. One process for identifying proteins in mixtures is to digest the mixture proteolytically and then acquire tandem mass spectra of peptides from each of the proteins in the mixture (Fig. 8).[129] Three enabling technologies have made this approach possible. ESI has permitted interfacing of liquid chromatography to resolve components of complex mixtures temporally prior to introduction into a tandem mass spectrometer.[102] 'On-the-fly' acquisition of tandem mass spectra through data-dependent control of the instrument has improved the efficiency of data acquisition.[103,130] Large numbers of tandem mass spectra (hundreds to thousands) can be automatically generated that would be impossible to acquire through manual operation of the instrument.[129] Lastly, direct and automated analysis of peptide tandem mass spectra is possible with the use of computer algorithms and

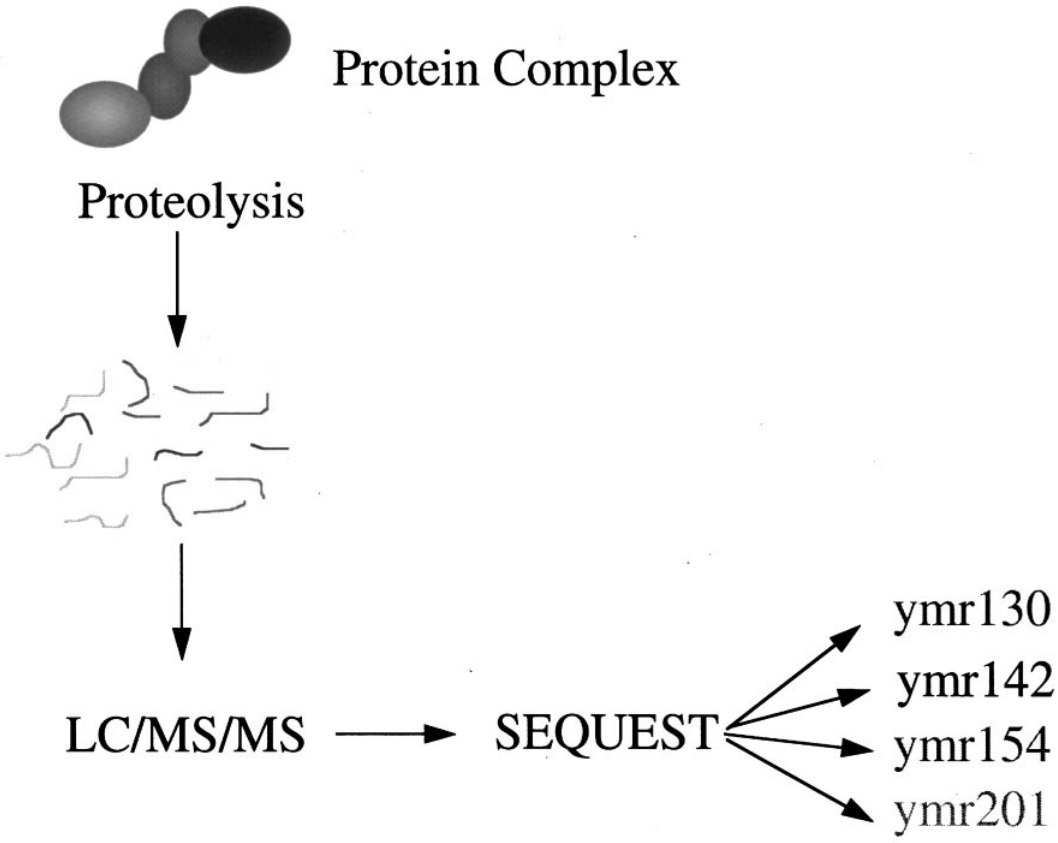

Figure 8. Shotgun identification of protein mixtures. A protein mixture is digested with a protease to produce a complicated mixture of peptides. The peptides are analyzed *en masse* using HPLC coupled to a tandem mass spectrometer. The tandem mass spectra are used to search the database using the SEQUEST database searching software and data review algorithms to identify the proteins present in the mixture.

databases. This approach is akin to a shotgun blast, creating many smaller fragments from a larger shell. The smaller fragments are analyzed and computer algorithms are then used to reconstruct the identities of the proteins. This approach is analogous to that used in large-scale genomic DNA sequencing.

Several reasons exist to extend the capability of the 'shotgun' identification approach for protein analysis. First, the end-product of many biological experiments, e.g. immunoprecipitation, is a small group of proteins (50–100, including background proteins). Traditionally, these proteins were separated by gel electrophoresis and then individually sequenced. By multiplexing the identification process through direct analysis of the digested collection of proteins, a rapid and automated method for protein identification can be achieved. Second, sample preparation and digestion procedures can be more aggressive with the addition of heating steps or chaotropic agents to aid complete proteolysis. Lastly, the potential for greater experimental throughput in allowing a wider range of physiological conditions to be examined is high. This type of approach is dependent on highly resolved separations integrated with tandem mass spectrometry and thus should become more powerful as the practice of two-dimensional liquid chromatography improves.[131]

Several applications of the shotgun identification method have been demonstrated. First and foremost, this method is at its most powerful in organisms with a completed genome sequence. When a genome is completed, the sequence should be available for every expressed protein. Strain differences may exist, but experience has shown this is not a complicating factor for protein identification using tandem mass spectrometry. The information not encoded by a genome sequence, at least not fully understood from the genome sequence information, but important to determine the function of a protein, is the subcellular location of a protein and protein–protein interactions. Some of this information can be hypothesized from the presence of sequence similarity to proteins of known function and cellular location or from the presence of sequence motifs implicating certain activity. However, protein activities and functions are dynamic with temporal and spatial components, and deciphering the information can only be accomplished through experimentation under a wide variety of conditions.

Protein localization

A general characteristic of protein function is the location in the cell where a protein is found. Nuclear proteins are likely to have very different functions from proteins found in the cytoplasm or in the plasma membrane. Determining the location of a protein is a good first step to narrowing the range of functions in which a

protein may participate. Cells contain a number of different compartments to segregate processes. For example, single-cell microorganisms are dependent upon constant communication with the environment outside the cell to obtain nutrients and to secrete waste products. *E. coli* and other Gram-negative bacteria possess an inner and outer membrane. The space located between the cell wall and the plasma membrane is called the periplasmic space and is a staging and storage area for materials entering and leaving the cell. In order to pass molecules into or out of the cell, they must cross the periplasmic space. Consequently, a set of soluble proteins are localized in this space to assist in the translocation of molecules into and out of the cell. To conduct a global survey of proteins that may be present in a specific location within a cell requires the use of biochemical methods to enrich collectively those proteins. Once the proteins of a compartment are enriched, they can be separated by one- or two-dimensional gel electrophoresis and identified. A new approach uses direct LC/MS/MS, without the aid of gel electrophoresis, to identify proteins from a subcellular location. Link *et al.*[132] enriched the proteins of the periplasmic space of *E. coli* and partially fractionated the proteins by using ion-exchange chromatography. After digesting the proteins in each FPLC fraction, the peptides were analyzed by using LC/MS/MS and database searching. This approach was used to identify a total of 80 proteins at a time when the *E. coli* genome was only 70–80% complete. A similar approach was taken by McCormack *et al.*[133] to characterize and identify peptides in the class II MHC antigen processing pathway.

Protein–protein interactions and protein complexes

Proteins involved in enzymatic or biochemical processes frequently form multi-protein complexes.[134] The proteins involved in processes or pathways can be identified by association under conditions indicating specific interaction. Biochemical techniques such as co-immunoprecipitation are widely used by biologists to identify interacting proteins (Fig. 9). Determination of protein–protein interactions has become an important component, particularly in the light of genome sequencing projects, to elucidate protein function or to implicate proteins in processes and pathways. Several biochemical strategies for determining protein–protein interactions have been adapted for use with the shotgun identification approach using LC/MS/MS and database searching.[129]

Protein interaction chromatography (PIC)

A sensitive method to detect protein–protein interactions binds a protein of interest to a solid support for use as 'bait.'[134] Cell extract is poured over the solid support and proteins with an affinity for the immobilized protein are retained on the column. The method can detect weakly associating proteins with binding constants in the range of 10^{-5} because the large abundance of protein immobilized forces the equilibrium. Proteins are then eluted from the column by disrupting the interactions with high salt solutions, chaotropic

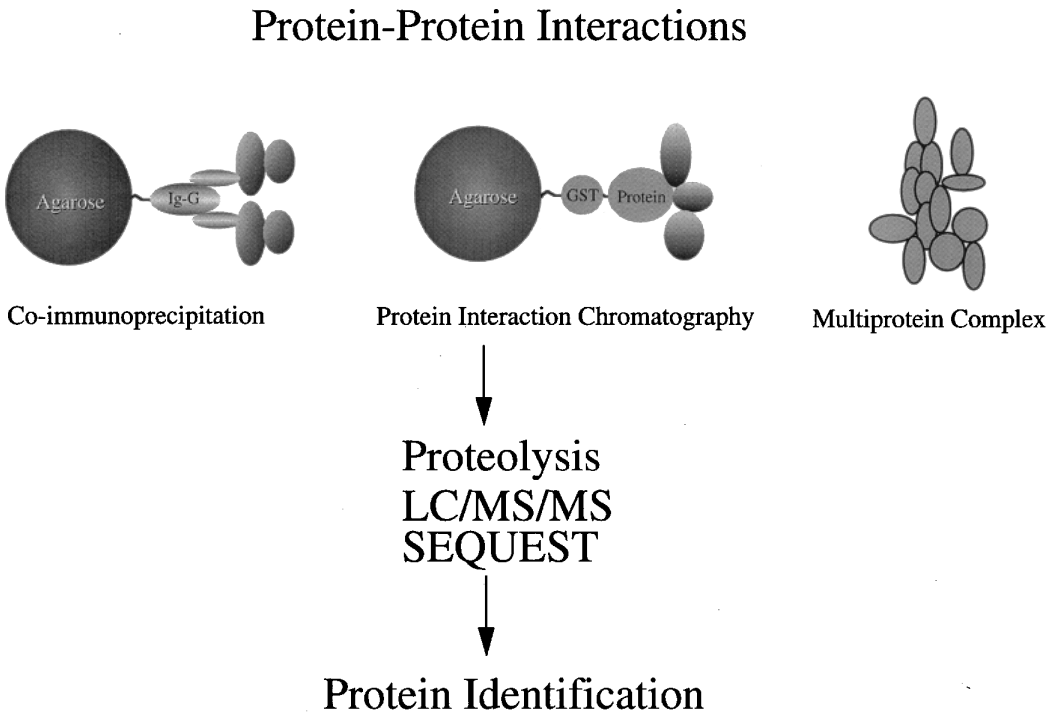

Figure 9. Three biochemical methods to determine protein–protein interactions. The first is co-immunoprecipitation. An antibody is used to precipitate a protein along with bound proteins. The second method, protein affinity interaction chromatography, uses a bound protein as 'bait' to bind to interacting proteins. The last method is purification of an intact protein complex.

Table 1. Summary of peptides and proteins identified that bound to the GST-SLA2 protein fusion column and eluted from the column in high salt conditions (0.6 M KCl)

Protein	M_r
NHPA non-histone chromosomal protein 6A	10784
PRE6 proteasome component	28421
G4P2 protein	29905
NFS1 nitrogen fixation-like protein	54475
DED1 putative ATP-dependent RNA helicase	65534
SSB1 heat shock protein	66452

agents or detergents. Tandem mass spectrometry can be used to identify directly the components enriched by PIC by digesting the proteins and analyzing with LC/MS/MS. A common strategy to immobilize proteins to a solid support is to create a gene fusion integrating a protein such as glutathionone-S-transferase (GST), a 12 kDa protein with a strong affinity for glutathione. The GST fusion can be readily bound to a solid support derivatized with glutathione. An example of proteins enriched and identified through the shotgun identification approach using a GST gene to the yeast protein Sla2p is shown in Table 1.[129]

Immunoprecipitation

A second classical method to identify interacting proteins is through the use of antibodies to coprecipitate a protein and any interacting proteins. Antibodies that specifically bind to the protein of interest must be available to perform the reaction. A common strategy incorporates a short amino acid sequence into a protein sequence to create an epitope to which antibodies are readily available. The antibody can be conjugated to a solid support to make recovery of the complex easier. Once bound, the solid support is washed with high-salt solutions, chaotropic agents or detergents to remove non-specifically bound proteins. If the enriched proteins are to be identified by LC/MS/MS analysis of the proteolysis products, the last washing step needs to be performed without detergent in the buffer. McCormack *et al.*[129] demonstrated the quantities of protein used for the direct LC/MS/MS strategy for protein identification was competitive with the quantities used for silver staining of the eluate on polyacrylamide gels. The proteins were derived from an immunoprecipitation of the *S. cerevisiae* RAS protein.

Table 2. Summary of proteins identified bound to microtubules

Protein	M_r
s22 40s ribosomal protein	14608
SSMla protein	24467
G4P2 protein suppressor protein	29905
L8300.8 hypothetical protein	33699
YM9718 hypothetical protein	34787
Heat shock protein	37572
SSA2 heat shock protein	69452
SSA1 heat shock protein	69749
L9576.2 protein	72535
YAT1 carnitine *o*-acetyltransferase	77263
eEF-2 translation elongation factor	93271
ARO1 pentafunctional aromatic polypeptide	174736

Protein complexes

Stable protein complexes are identified as groups of proteins that co-purify under a variety of conditions. Traditional biochemical methods such as gel filtration chromatography, density gradient centrifugation and ion-exchange chromatography are used to purify protein complexes. Stable protein complexes such as the ribosomal complex or structural complexes such as the microtubule constitute large groups of proteins. In many cases, other proteins beyond the core proteins of the complex interact to control or regulate the activity of the complex. To enrich for these other proteins, the core protein complex can be used. To study microtubule formation and regulation in *S. cerevisiae*, bovine tubulin is used to form microtubules since it can be obtained in much greater abundance than tubulin from yeast. Monomeric tubulin can be induced to polymerize and form microtubules by the addition of taxol. The macromolecular complex has been used to enrich proteins in *S. cerevisiae* with a specific affinity for the complex. By varying the conditions of incubation, e.g. buffers and co-factors, different sets of proteins can be enriched.[135] A yeast whole cell lysate is added to the macromolecular complex and washed with different salt conditions. The proteins that elute under the different conditions are collected and digested and their identities determined by LC/MS/MS and database searching. A list of proteins identified as bound to polymerized microtubules is given in Table 2.

CONCLUSION

Mass spectrometry has reached new levels of sensitivity and throughput for protein analysis, concurrent with the biological sciences entering a new era of discovery. As biologists begin the process of gene functional analysis, several issues are becoming clear.[136] Computational strategies (bioinformatics) can provide a good indication of likely gene function by identifying similar sequence features among related genes and proteins of different organisms, but cannot flesh out the details of pathways or physiological processes. Additionally, proteins are pleiotropic, that is, they will be functionally required in different places, or at different times. A consequence of pleiotrophy is that protein function will have spatial, temporal and tissue specificity. Protein function will, therefore, be complex, requiring systematic functional studies. Such studies will benefit from the ability to identify and quantitate gene transcripts, proteins and covalent modifications to proteins. Mass spec-

trometry will clearly assume a role in protein and protein modification identification as part of functional studies and the continued improvement of mass spectrometry-based approaches will create an even more valuable tool for the emerging fields of functional genomics and proteomics.

Acknowledgements

The author acknowledges the financial support provided by grants from the National Institutes of Health (GM52095) and the National Science Foundation, Science and Technology Center (BIR9214821). Dr Ashok Dongre is thanked for helpful comments.

REFERENCES

1. R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith and J. C. Venter, *Science* **269**, 496 (1995).
2. J. F. Tomb, O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzegerald, N. Lee, M. D. Adams, J. C. Venter, *et al*., *Nature* (*London*) **388**, 539 (1997).
3. R. Himmelrich, H. Hilbert, H. Plagens, E. Pirkl, B. C. Li and R. Herrmann, *Nucleic Acids Res*. **24**, 4420 (1996).
4. C. J. Bult, O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. M. Geoghagen and J. C. Venter, *Science* **273**, 1058 (1996).
5. C. M. Fraser, J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, *et al*., *Science* **270**, 397 (1995).
6. F. R. Blattner, G. r. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau and Y. Shao, *Science* **277**, 1453 (1997).
7. T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda and S. Tabata, *DNA Res*. **3**, 109 (1996).
8. K. W. Minton, *Mol. Microbiol*. **13**, 9 (1994).
9. R. Ciulla, C. Clougherty, N. Belay, S. Krishnan, C. Zhou, D. Bryd and M. F. Roberts, *J. Bacteriol*. **176**, 3177 (1994).
10. M. Schena, D. Shalon, R. W. Davis and P. O. Brown, *Science* **270**, 467 (1995).
11. D. Shalon, S. J. Smith and P. O. Brown, *Genome Res*. **6**, 639 (1996).
12. J. R. Yates, III and A. L. McCormack, *Anal. Chem*. **68**, 534A (1996).
13. K. L. Busch, *J. Mass Spectrom*. **30**, 233 (1995).
14. S. J. Gaskell, *J. Mass Spectrom*. **32**, 677 (1996).
15. H. R. Morris, A. Dell and R. A. McDowell, *Biomed. Mass Spectrom*. **8**, 463 (1981).
16. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse, *Science* **246**, 64 (1989).
17. F. Hillenkamp, M. Karas, R. C. Beavis and B. T. Chait, *Anal. Chem*. **63**, 1193A (1991).
18. J. G. Boyle and C. M. Whitehouse, *Anal. Chem*. **64**, 2084 (1992).
19. K. D. Henry, E. R. Williams, B. H. Wang, F. W. McLafferty, J. Shabanowitz and D. F. Hunt, *Proc. Natl. Acad. Sci. USA* **86**, 9075 (1989).
20. S. A. McLuckey, G. J. Van Berkel, D. E. Goeringer and G. L. Glish, *Anal. Chem*. **66**, 737A (1994).
21. K. Jonscher, G. Currie, A. L. McCormack and J. R. Yates, *Rapid Commun. Mass Spectrom*. **7**, 20 (1993).
22. J. A. Castoro and C. L. Wilkins, *Anal. Chem*. **65**, 2621 (1993).
23. E. D. Lee, W. Muck, J. D. Henion and T. R. Covey, *J. Chromatogr*. **458**, 313 (1988).
24. T. R. Covey, E. C. Huang and J. D. Henion, *Anal. Chem*. **63**, 1193 (1991).
25. M. Karas and F. Hillenkamp, *Anal. Chem*. **60**, 2299.
26. M. Karas and F. Hillenkamp, *AIP Conf. Proc*., No. 288, 447 (1993).
27. D. H. Russell, M. A. White and T. Solouki, *Org. Mass Spectrom*. **27**, 827 (1992).
28. R. C. Beavis and B. T. Chait, *Anal. Chem*. **62**, 1836 (1990).
29. R. C. Beavis and B. T. Chait, *Proc. Natl. Acad. Sci. USA* **87**, 6873 (1990).
30. T. D. Wood, L. H. Chen, C. B. White, P. C. Babbitt, G. L. Kenyon and F. W. McLafferty, *Proc. Natl. Acad. Sci. USA* **93**, 12051 (1996).
31. J. R. Yates, *Methods Enzymol*. **271**, 351 (1996).
32. F. W. McLafferty, *Tandem Mass Spectrometry*. Wiley–Interscience, New York (1983).
33. A. R. Dongr'e, A. Somogyi and V. H. Wysocki, *J. Mass Spectrom*. **31**, 339 (1996).
34. J. C. Schwartz, A. P. Wade, C. G. Enke and R. G. Cooks, *Anal. Chem*. **62**, 1809 (1990).
35. R. E. March, *J. Mass Spectrom*. **32**, 351 (1996).
36. J. N. Louris, J. W. Amy, T. Y. Ridley and R. G. Cooks, *Int. J. Mass Spectrom. Ion Processes* **88**, 2 (1989).
37. E. Hoffmann, *J. Mass Spectrom*. **31**, 129 (1996).
38. J. N. Louris, J. S. Brodbelt Lustig, R. G. Cooks, G. L. Glish, G. J. van Berkel and S. A. McLuckey, *Int. J. Mass Spectrom. Ion Processes* **96**, 117 (1990).
39. S. A. McLuckey, G. L. Glish and G. J. Van Berkel, *Int. J. Mass Spectrom. Ion Processes* **106**, 213 (1991).
40. R. K. Julian, Jr and R. G. Cooks, *Anal. Chem*. **65**, 1827 (1993).
41. S. A. McLuckey, D. E. Goeringer and G. L. Glish, *Anal. Chem*. **64**, 1455 (1992).
42. M. Guilhaus, *J. Mass Spectrom*. **30**, 1519 (1995).
43. R. C. Beavis and B. T. Chait, *Chem. Phys. Lett*. **181**, 479 (1991).
44. R. S. Brown and J. J. Lennon, *Anal. Chem*. **67**, 1998 (1995).
45. A. N. Verentchikov, W. Ens and K. G. Standing, *Anal. Chem*. **66**, 126 (1994).
46. H. R. Morris, T. Paxton, A. Dell, J. Langhorne, M. Berg, R. S. Bordoli, J. Hoyes and R. H. Bateman, *Rapid Commun. Mass Spectrom*. **10**, 889 (1996).
47. B. Spengler, *J. Mass Spectrom*. **32**, 1017 (1996).
48. B. Spengler, D. Kirsch, R. Kaufmann and E. Jaeger, *Rapid Commun. Mass Spectrom*. **6**, 105 (1992).
49. R. S. Brown and J. J. Lennon, *Anal. Chem*. **67**, 3990 (1995).
50. Y. H. Liu, J. Bai, X. Liang, D. M. Lubman and P. J. Venta, *Anal. Chem*. **67**, 3482 (1995).
51. T. Houthaeve, H. Gausepohl, M. Mann and K. Ashman, *FEBS Lett*. **376**, 91 (1995).
52. R. T. Kennedy and J. W. Jorgenson, *Anal. Chem*. **61**, 1128 (1989).
53. C. Flurer, C. Borra, S. Beale and M. Novotny, *Anal. Chem*. **60**, 1826 (1988).
54. M. T. Davis, D. C. Stahl and T. D. Lee, *J. Am. Soc. Mass*

Spectrom. **5**, 571 (1995).
55. M. T. Davis, D. C. Stahl, S. A. Hefta and T. D. Lee, *Anal. Chem*. **67**, 4549 (1995).
56. M. T. Davis and T. D. Lee, *J. Am. Soc. Mass Spectrom*. **8**, 1059 (1997).
57. J. W. Jorgenson and K. D. Lukacs, *Science* **222**, 266 (1983).
58. J. H. Wahl, D. R. Goodlett, H. R. Udseth and R. D. Smith, *Electrophoresis* **14**, 448 (1993).
59. P. Thibault, C. Paris and S. Pleasance, *Rapid Commun. Mass Spectrom*. **5**, 484 (1991).
60. M. A. Moseley, J. W. Jorgenson, J. Shabonowitz, D. F. Hunt and K. B. Tomer, *J. Am. Soc. Mass Spectrom*. **3**, 289 (1992).
61. A. J. Tomlinson, L. M. Benson, W. D. Braddock, R. P. Oda and S. Naylor, *J. High Resolut. Chromatogr*. **17**, 729 (1994).
62. A. J. Tomlinson, L. M. Benson and S. Naylor, *J. Capillary Electrophor*. **1**, 127 (1994).
63. M. A. Strausbauch, B. J. Madden, P. J. Wettstein and J. P. Landers, *Electrophoresis* **16**, 541 (1995).
64. A. J. Tomlinson, L. M. Benson, S. Jameson and S. Naylor, *Electrophoresis* **17**, 1801 (1996).
65. S. Naylor and A. J. Tomlinson, *Biomed. Chromatogr*. **10**, 325 (1996).
66. D. Figeys, A. Ducret, J. R. Yates and R. Aebersold, *Nature Biotechnol*. **14**, 1579 (1996).
67. D. C. Gale and R. D. Smith, *Rapid Commun. Mass Spectrom*. **7**, 1017 (1993).
68. M. R. Emmett and R. M. Caprioli, *J. Am. Soc. Mass Spectrom*. **5**, 605 (1994).
69. M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis and M. Mann, *Nature* (*London*) **379**, 466 (1996).
70. M. S. Wilm and M. Mann, *Int. J. Mass Spectrom. Ion Processes* **136**, 2 (1994).
71. M. Wilm, G. Neubauer and M. Mann, *Anal. Chem*. **68**, 527 (1996).
72. H. R. Morris, M. Panico and G. W. Taylor, *Biochem. Biophys. Res. Commun*. **117**, 299 (1983).
73. B. W. Gibson and K. Biemann, *Proc. Natl. Acad. Sci. USA* **81**, 1956 (1984).
74. C. Eckerskorn, K. Strupat, M. Karas, F. Hillenkamp and F. Lottspeich, *Electrophoresis* **13**, 664 (1992).
75. M. M. Vestling and C. Fenselau, *Biochem. Soc. Trans*. **22**, 547 (1994).
76. R. R. Ogorzalek Loo, C. Mitchell, T. I. Stevenson, S. A. Martin, W. M. Hines, P. Juhasz, D. H. Patterson, J. M. Peltier, J. A. Loo and P. C. Andrews, *Electrophoresis* **18**, 382 (1997).
77. C. Eckerskorn, K. Strupat, D. Schleuder, D. Hochstrasser, J. C. Sanchez, F. Lottspeich and F. Hillenkamp, *Anal. Chem*. **69**, 2888 (1997).
78. D. Fabris, M. M. Vestling, M. M. Cordero, V. M. Doroshenko, R. J. Cotter and C. Fenselau, *Rapid Commun. Mass Spectrom*. **9**, 1051 (1995).
79. C. V. Bradley, D. H. Williams and M. R. Hanley, *Biochem. Biophys. Res. Commun*. **104**, 1223 (1982).
80. D. H. Patterson, G. E. Tarr, F. E. Regnier and S. A. Martin, *Anal. Chem*. **67**, 3971 (1995).
81. A. S. Woods, A. Y. Huang, R. J. Cotter, G. R. Pasternack, D. M. Pardoll and E. M. Jaffee, *Anal. Biochem*. **226**, 15 (1995).
82. B. T. Chait, R. Wang, R. C. Beavis and S. B. Kent, *Science* **262**, 89 (1993).
83. M. Bartlet Jones, W. A. Jeffery, H. F. Hansen and D. J. Pappin, *Rapid Commun. Mass Spectrom*. **8**, 737 (1994).
84. W. J. Henzel, T. M. Billeci, J. T. Stults, S. C. Wong, C. Grimley and C. Watanabe, *Protein Sci*. **2**, Suppl. 1, 90 (1993).
85. D. J. C. Pappin, A. J. Bleasby, C. W. Sutton and J. S. Cottrell, *Protein Sci*. **2**, Suppl. 1, 90 (1993).
86. J. R. Yates, S. Speicher, P. R. Griffin and T. Hunkapiller, *Anal. Biochem*. **214**, 397 (1993).
87. P. James, M. Quadroni, E. Carafoli and G. Gonnet, *Biochem. Biophys. Res. Commun*. **195**, 58 (1993).
88. M. Mann, P. Hojrup and P. Roepstorff, *Biol. Mass Spectrom*. **22**, 338 (1993).
89. P. James, M. Quadroni, E. Carafoli and G. Gonnet, *Protein Sci*. **3**, 1347 (1994).
90. K. R. Clauser, S. C. Hall, D. M. Smith, J. W. Webb, L. E. Andrews, H. M. Tran, L. B. Epstein and A. L. Burlingame, *Proc. Natl. Acad. Sci. USA* **92**, 5072 (1995).
91. S. D. Patterson, D. Thomas and R. A. Bradshaw, *Electrophoresis* **17**, 877 (1996).
92. E. J. Takach, W. M. Hines, D. H. Patterson, P. Juhasz, A. M. Falick, M. L. Vestal and S. A. Martin, *J. Protein Chem*. **16**, 363 (1997).
93. O. N. Jensen, A. Podtelejnikov and M. Mann, *Rapid Commun. Mass Spectrom*. **10**, 1371 (1996).
94. S. J. Cordwell, V. C. Wasinger, A. Cerpa Poljak, M. W. Duncan and I. Humphery Smith, *J. Mass Spectrom*. **32**, 370 (1997).
95. S. J. Cordwell and I. Humphrey-Smith, *Electrophoresis* **18**, 1410 (1997).
96. D. F. Hunt, A. M. Buko, J. M. Ballard, J. Shabanowitz and A. B. Giordani, *Biomed. Mass Spectrom*. **8**, 397 (1981).
97. D. F. Hunt, J. R. Yates, III, J. Shabanowitz, S. Winston and C. R. Hauer, *Proc. Natl. Acad. Sci. USA* **83**, 6233 (1986).
98. K. Biemann, *Methods Enzymol* **193**, 455 (1990).
99. P. Roepstorff and J. Fohlmann, *Biomed. Mass Spectrom*. **11**, 601 (1984).
100. A. L. McCormack, A. Somogyi, A. R. Dongr'e and V. H. Wysocki, *Anal. Chem*. **65**, 2859 (1993).
101. J. R. Yates and J. K. Eng, *US Pat*. 5 538 897 (1996).
102. J. K. Eng, A. L. McCormack and J. R. Yates, III, *J. Am. Soc. Mass Spectrom*. **5**, 976 (1994).
103. J. R. Yates, J. K. Eng, A. L. McCormack and D. Schieltz, *Anal. Chem*. **67**, 1426 (1995).
104. J. R. Yates, III, J. K. Eng, A. L. McCormack and D. Schieltz, *Anal. Chem*. **67**, 1426 (1995).
105. J. R. Yates, III, J. K. Eng and A. L. McCormack, *Anal. Chem*. **67**, 3202 (1995).
106. M. Mann and M. Wilm, *Anal. Chem*. **66**, 4390 (1994).
107. P. R. Griffin, M. J. MacCoss, J. K. Eng, R. A. Blevins, J. S. Aaronson and J. R. Yates, *Rapid Commun. Mass Spectrom*. **9**, 1546 (1995).
108. K. R. Jonscher and J. R. Yates, *Anal. Biochem*. **244**, 1 (1997).
109. J. R. Yates, J. K. Eng, C. Klausner and A. L. Burlingame, *J. Am. Soc. Mass Spectrom*. **7**, 1089 (1996).
110. K. Gevaert, J. L. Verschelde, M. Puype, J. Van Damme, M. Goethals, S. De Boeck and J. Vandekerckhove, *Electrophoresis* **17**, 918 (1996).
111. K. Gevaert, H. De Mol, J. L. Verschelde, J. Van Damme, S. De Boeck and J. Vandekerckhove, *J. Protein Chem*. **16**, 335 (1997).
112. J. E. Fitton, D. F. Hunt, J. Marasco, J. Shabanowitz, S. Winston and A. Dell, *FEBS Lett*. **169**, 25 (1984).
113. D. F. Hunt, J. R. Yates, J. Shabanowitz, N. Z. Zhu, T. Zirino, B. A. Averill, S. T. Daurat Larroque, J. G. Shewale, R. M. Roberts and K. Brew, *Biochem. Biophys. Res. Commun*. **114**, 1154 (1987).
114. D. F. Hunt, J. R. Yates, J. Shabanowitz, S. Winston and C. R. Hauer, *Proc. Natl. Acad. Sci. USA* **83**, 6233 (1986).
115. A. L. Cox, J. Skipper, Y. Chen, R. A. Henderson, T. L. Darrow, J. Shabanowitz, V. H. Engelhard, D. F. Hunt and C. L. Slingluff, Jr, *Science* **264**, 716 (1994).
116. D. F. Hunt, R. A. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A. L. Cox, E. Appella and V. H. Engelhard, *Science* **225**, 1261 (1992).
117. D. F. Hunt, H. Michel, T. A. Dickinson, J. Shabanowitz, A. L. Cox, K. Sakaguchi, E. Appella, H. M. Grey and A. Sette, *Science* **256**, 1817 (1992).
118. M. Witty, A. D. Wallace Cook, H. Albrecht, A. J. Spano, H. Michel, J. Shabanowitz, D. F. Hunt, M. P. Timko and A. G. Smith, *Plant Physiol*. **103**, 139 (1993).
119. Y. Y. Wo, A. L. McCormack, J. Shabanowitz, D. F. Hunt, J. P. Davis, G. L. Mitchell and R. L. Van Etten, *J. Biol. Chem*. **267**, 10856 (1992).
120. J. Lingner, T. R. Hughes, A. Shevchenko, M. Mann, V. Lundblad and T. R. Cech, *Science* **276**, 561 (1997).
121. J. Vandekerckhove, G. Bauw, M. Puype, J. Van Damme and M. Van Montagu, *Eur. J. Biochem*. **152**, 9 (1985).
122. R. H. Aebersold, D. B. Teplow, L. E. Hood and S. B. Kent, *J. Biol. Chem*. **261**, 4229 (1986).
123. R. H. Aebersold, J. Leavitt, R. A. Saavedra, L. E. Hood and S. B. Kent, *Proc. Natl. Acad. Sci. USA* **84**, 6970 (1987).
124. U. Hellman, C. Wernstedt, J. G'Onez and C. H. Heldin, *Anal.*

Biochem. **224**, 451 (1995).
125. A. Shevchenko, M. Wilm, O. Vorm and M. Mann, *Anal. Chem*. **68**, 850 (1996).
126. C. W. Sutton, K. S. Pemberton, J. S. Cottrell, J. M. Corbett, C. H. Wheeler, M. J. Dunn and D. J. Pappin, *Electrophoresis* **16**, 308 (1995).
127. A. Shevchenko, O. N. Jensen, A. V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie and M. Mann, *Proc. Natl. Acad. Sci. USA* **93**, 14440 (1996).
128. A. J. Link, L. G. Hays, E. B. Carmack and J. R. Yates, *Electrophoresis* **18**, 1314 (1997).
129. A. L. McCormack, D. M. Schieltz, B. Goode, S. Yang, G. Barnes, D. Drubin and J. R. Yates, *Anal. Chem*. **69**, 767 (1997).
130. D. C. Stahl, K. M. Swiderek, M. T. Davis and T. D. Lee, *J. Am. Soc. Mass Spectrom*. **7**, 532 (1996).
131. J. P. Larmann, Jr, A. V. Lemmo, A. W. Moore, Jr and J. W. Jorgenson, *Electrophoresis* **14**, 439 (1993).
132. A. J. Link, E. Carmack and J. R. Yates, *Int. J. Mass Spectrom. Ion Processes* **160**, 303 (1997).
133. A. L. McCormack, J. K. Eng, P. C. DeRoos, A. Y. Rudensky and J. R. Yates, in *Biochemical and Biotechnological Applications of Electrospray Ionization in Mass Spectrometry*, edited by A. P. Synder, p. 207. American Chemical Society, Washington, DC (1995).
134. E. M. Phizicky and S. Fields, *Microbiol. Rev*. **59**, 94 (1995).
135. G. Barnes, K. A. Louie and D. Botstein, *Mol. Biol. Cell* **3**, 29 (1992).
136. G. L. Miklos and G. M. Rubin, *Cell* **86**, 521 (1996).