Annu. Rev. Biochem. 2001. 70:437–73

ANALYSIS OF PROTEINS AND PROTEOMES BY MASS SPECTROMETRY

Matthias Mann,[1,2] Ronald C. Hendrickson,[2] and Akhilesh Pandey[3]

[1]*Protein Interaction Laboratory and Center for Experimental BioInformatics (CEBI), Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark; e-mail: mann@bmb.sdu.dk*

[2]*MDS Proteomics A/S, Staermosegaardsvej 6, 5230 Odense M, Denmark; e-mail: hendrickson@protana.com*

[3]*Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, and Brigham and Women's Hospital, Boston, Massachusetts 02115; e-mail: pandey@bmb.sdu.dk*

Key Words functional genomics, signal transduction, posttranslational modifications, protein interaction, phosphorylation, proteomics

■ **Abstract** A decade after the discovery of electrospray and matrix-assisted laser desorption ionization (MALDI), methods that finally allowed gentle ionization of large biomolecules, mass spectrometry has become a powerful tool in protein analysis and the key technology in the emerging field of proteomics. The success of mass spectrometry is driven both by innovative instrumentation designs, especially those operating on the time-of-flight or ion-trapping principles, and by large-scale biochemical strategies, which use mass spectrometry to detect the isolated proteins. Any human protein can now be identified directly from genome databases on the basis of minimal data derived by mass spectrometry. As has already happened in genomics, increased automation of sample handling, analysis, and the interpretation of results will generate an avalanche of qualitative and quantitative proteomic data. Protein-protein interactions can be analyzed directly by precipitation of a tagged bait followed by mass spectrometric identification of its binding partners. By these and similar strategies, entire protein complexes, signaling pathways, and whole organelles are being characterized. Posttranslational modifications remain difficult to analyze but are starting to yield to generic strategies.

CONTENTS

INTRODUCTION

Mass spectrometry (MS) is a venerable technique whose beginnings date back to the early days of the last century. Among the analytical techniques, MS holds a special place because it measures an intrinsic property of a molecule, its mass, with very high sensitivity and therefore it is used in an amazingly wide range of applications. Beginning in the 1980s and on a larger scale in the 1990s, mass spectrometry has played an increasingly significant role in the biological sciences. Why has it taken so long? Mainly because mass spectrometers require charged, gaseous molecules for analysis. Biomolecules being large and polar, however, they are not easily transferred into the gas phase and ionized. Electrospray (ES) (1) and matrix-assisted laser desorption ionization (MALDI) (2) are the ionization techniques that should be credited most for the success of mass spectrometry in the life sciences. These methods were developed in the late 1980s and were the basis for the increasingly powerful instrumentation that became available a few

years later. Major advances were also made in sample preparation for MS, a crucial area for overall feasibility and sensitivity of analysis. Starting in 1993, software algorithms were published that allowed the correlation of mass spectrometric data obtained for a protein with the increasingly populated sequence databases. In retrospect, this event marked the transformation of mass spectrometry into a large-scale, functional genomics technique. The last few years have seen development of even more powerful instrumentation and algorithms for protein characterization, a trend that shows no signs of slowing down.

At the same time MS was being developed to meet the demands of molecular biology for high sensitivity, the concept of proteomics began to be popularized. Proteomics is now defined as the large-scale analysis of the function of genes and is becoming a central field in functional genomics (3). The major tool to study purified proteins in this field is mass spectrometry. The history of proteomics dates back to the discovery of two-dimensional gels in the 1970s, which provided the first feasible way of displaying hundreds or thousands of proteins on a single gel (4, 5). Identification of the spots separated on these gels remained laborious and was limited to the most abundant proteins until the 1990s, when biological mass spectrometry had developed into a sufficiently sensitive and robust technique. Today, mass spectrometry is an essential element of the proteomics field. Indeed, researchers are now successfully harnessing the power of MS to supersede the two-dimensional gels that originally gave proteomics its impetus.

Currently, the uses of MS in proteomics are in three major areas. MS is the preferred technique for characterization and quality control of recombinant proteins and other macromolecules, an important task in the field of biotechnology. It is also commonly used for protein identification, either in classical biochemical projects or in large-scale proteomic ones. Finally, because MS measures the molecular weight of a protein, it is the method of choice for the detection and characterization of posttranslational modifications and potentially can identify any covalent modification that alters the mass of a protein.

In this review we describe the principles of the MS ionization methods and the major types of instruments currently in use and under development, discuss strategies for the analysis of intact proteins and for the sequencing of peptides obtained after enzymatic degradation of the protein, and cover the identification of proteins in databases by MS data. Analysis of posttranslational modifications and the recent development of quantitative mass spectrometry using stable isotopes are also discussed.

Of equal importance to the technological developments is the development of strategies for applying MS to solve real and interesting biological problems. We describe the difference between expression proteomics, which aims to measure up- and down-regulation of protein levels, and functional proteomics as exemplified in several recent strategies aimed at the characterization of cellular compartments, multiprotein complexes, and signaling pathways.

IONIZATION METHODS

MALDI

Matrix-assisted laser desorption ionization (MALDI), developed by Karas & Hillenkamp in the late 1980s (2), is one of the two "soft ionization "methods. To generate gas phase, protonated molecules, a large excess of matrix material is coprecipitated with analyte molecules (that is, the molecules to be analyzed) by pipetting a submicroliter volume of the mixture onto a metal substrate and allowing it to dry. The resulting solid is then irradiated by nanosecond laser pulses, usually from small nitrogen lasers with a wavelength of 337 nm. The matrix is typically a small organic molecule with absorbance at the wavelength of the laser employed. Work with biomolecules almost exclusively uses matrices of α-cyano-4-hydroxycinnamic acid or dihydrobenzoic acid (DHB). Matrices differ in the amount of energy they impart to the biomolecules during desorption and ionization and hence the degree of fragmentation (unimolecular decay) that they cause. The α-cyano matrix, which generally leads to the highest sensitivity in MALDI, is "hotter" than DHB, so the latter is preferred when the ions need to be stable for milliseconds in trapping experiments rather than microseconds in time-of-flight experiments (see below). Numerous variations on the basic dried-droplet method of sample preparation for MALDI have been described. For example, the matrix can be laid down in a microcrystalline thin film on the substrate, which leads to better adherence so that the sample can be washed more vigorously and which also presents a larger crystalline surface from which the ions can be desorbed (6). Admixture with the matrix can be beneficial; for example, nitrocellulose increases the even representation of peptides (7), and fructose can lower the energy state of the desorbed ions.

The precise nature of the ionization process in MALDI is still largely unknown and the signal intensities depend on incorporation of the peptides into crystals, their likelihood of capturing and/or retaining a proton during the desorption process, and a number of other factors including suppression effects in peptide mixtures. For example, peptides with a C-terminal arginine generally result in higher signals than peptides with a C-terminal lysine (8). For these reasons it is difficult to relate peptide peak height with the quantity of sample present unless an internal standard is used.

The mass range below 500 daltons (Da) is often obscured by matrix-related ions in MALDI. Proteins generally undergo fragmentation to some extent during MALDI, resulting in broad peaks and loss in sensitivity; therefore MALDI is mostly applied to the analysis of peptides.

Electrospray

Electrospray mass spectrometry (ESMS) has been developed for use in biological mass spectrometry by Fenn et al (1). Liquid containing the analyte is

pumped at low microliter-per-minute flow rates through a hypodermic needle at high voltage to electrostatically disperse, or electrospray, small, micrometer-sized droplets, which rapidly evaporate and which impart their charge onto the analyte molecules. This ionization process takes place in atmosphere and is therefore very gentle (without fragmentation of analyte ions in the gas phase). The molecules are transferred into the mass spectrometer with high efficiency for analysis. To stabilize the spray, a nebulizer gas or some other device is often employed.

A wide range of compounds can be analyzed by ESMS; the only requirement is that the molecule be sufficiently polar to allow attachment of a charge. This includes proteins, oligonucleotides, sugars (with less sensitivity, as sodium rather than hydrogen is the charging agent), and polar lipids. For a given compound, the signal strength (peak height in the spectrum) increases linearly with the analyte concentration over a wide range until saturation occurs. However, the signal is to a first approximation independent of liquid flow rate, which makes it desirable to operate at the lowest flow rate possible. There does not seem to be an upper mass limit to analysis by ESMS. Large ions are typically multiply charged (proteins and peptides by added protons in the positive mode and abstracted protons in the negative mode), which brings them into the range of mass-to-charge (*m/z*) ratios of typical mass spectrometers (note that a mass spectrometer measures *m/z* rather than the mass). The distribution of charges gives rise to the typical multiple charge envelope (see Figure 3 below). These spectra can be simplified by deconvolution, an algorithm that sums up the signal intensity into a single peak at the molecular weight of the analyte. Very complex mixtures can be analyzed by ESMS, but the spectra become increasingly difficult to interpret as the molecular weight of the components and their number increases.

Electrospray is typically performed in either the infusion mode, the nanoelectrospray format, or in combination with high-performance liquid chromatography. In the infusion mode, the sample is simply introduced into a continuous liquid stream (typically a mixture of organic and aqueous liquid such as 50:50 MeOH:H_2O) via an injection valve. Flow rates are usually between 0.5 and several microliters per minute. Samples have to be substantially free of salt and detergent, but can conveniently be cleaned up in a reversed-phase packing loop in the injector valve. Nanoelectrospray (9, 10) is a miniaturized version of electrospray that operates without pumps and at very low flow rates in the range of a few nanoliters per minute. It is performed in pulled glass capillaries with an inner diameter at the tip of about one micrometer. A microliter volume of sample can be analyzed for more than an hour at full signal strength, which allows complex sequencing experiments to be performed (see below). When liquid chromatography and mass spectrometry are coupled (LC-MS), MS analysis of the components of the sample takes place on-line as they elute from the chromatography column. In this scenario, sample cleanup, separation, and concentration are all achieved in a single step.

MASS SPECTROMETERS

The last years have seen tremendous improvements in mass spectrometric instrumentation. It is interesting to note that the cumulative advances in MS instrumentation have occurred at a much faster pace than the oft-cited increases in computer processing speeds. This progress is the result both of incremental advances in existing mass spectrometers and of entirely novel concepts and novel combinations of existing instruments. Essentially, all mass spectrometers measure the mass-to-charge ratio of analytes such as proteins, peptides, or peptide fragments. Three different principles may be applied to achieve mass separation: separation on the basis of time-of-flight (TOF MS), separation by quadrupole electric fields generated by metal rods (quadrupole MS), or separation by selective ejection of ions from a three-dimensional trapping field (ion trap or Fourier transform ion cyclotron MS). For structural analysis, such as peptide sequencing, two steps of mass spectrometry are performed in tandem (tandem mass spectrometry or MS/MS), which can be done by employing the same separation principle twice or by combining two different MS separation principles. Both MALDI and electrospray can be coupled to any one of these three methods of separation. The fact that MALDI produces short bursts of ions in the vacuum and electrospray produces a continuous beam of ions in atmosphere typically led to coupling of MALDI with TOF MS and electrospray with quadrupole and ion-trapping MS.

MALDI Time-of-Flight Mass Spectrometer

Figure 1 shows a schematic of the instrument combining MALDI and TOF in a mass spectrometer. Samples are deposited on a metal substrate capable of holding between one and several hundred analyte spots. These spots are then irradiated by a laser pulse, to generate a short burst of ions. The ions are accelerated to a fixed amount of kinetic energy and travel down a flight tube. The small ions have a higher velocity and are recorded on a detector before the larger ones, producing the time-of-flight (TOF) spectrum.

Several dozen to hundreds of laser shots are averaged to produce the final MALDI spectrum. Performance in modern reflector MALDI mass spectrometers is typically in the range of a few parts per million in mass accuracy, and only about a femtomole of peptide material needs to be deposited on the MALDI target to produce a signal.

The fact that some of the peptide ions decay because of the energy departed in the desorption process has been used to obtain structural information in a technique termed postsource decay (11). It can provide important structural information but does not serve as a general peptide sequencing method because it is not sufficiently sensitive and simple to control.

Electrospray Quadrupole Mass Spectrometers

Electrospray is most often combined with a quadrupole mass spectrometer. The quadrupole is a mass filter, which consists of four rods to which an oscillating

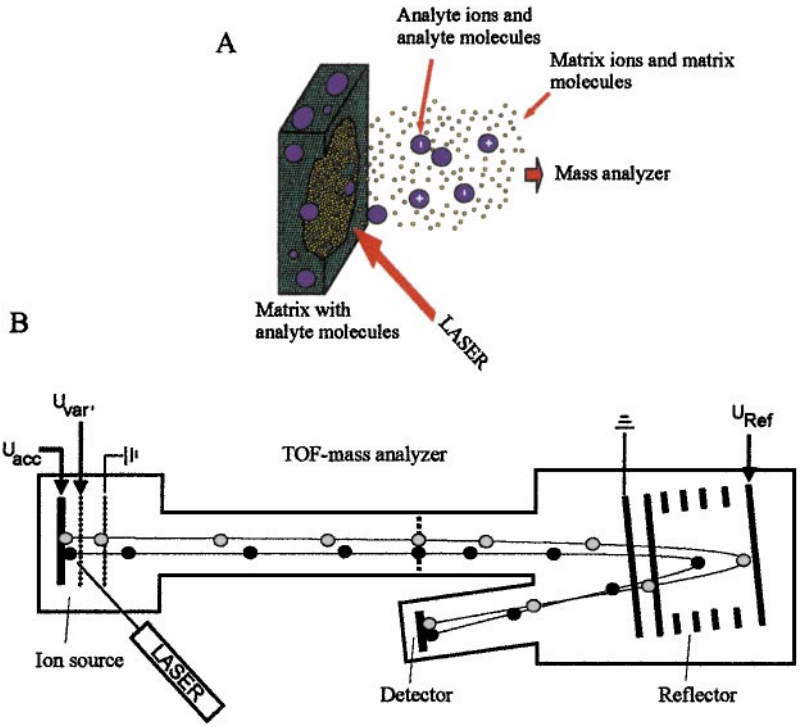


Figure 1 Schematic of MALDI process and instrument. (*A*) A sample cocrystallized with the matrix is irradiated by a laser beam, leading to sublimation and ionization of peptides. (*B*) About 100–500 ns after the laser pulse, a strong acceleration field is switched on (delayed extraction), which imparts a fixed kinetic energy to the ions produced by the MALDI process. These ions travel down a flight tube and are turned around in an ion mirror, or reflector, to correct for initial energy differences. The mass-to-charge ratio is related to the time it takes an ion to reach the detector; the lighter ions arrive first. The ions are detected by a channeltron electron multiplier.

electric field is applied and which lets only a certain mass pass through (the other masses are on unstable trajectories and do not reach the detector). By scanning the amplitude of the electric field and recording the ions at the detector, one obtains a mass spectrum. To date, most peptide sequencing experiments have been performed on triple quadrupole instruments that consist of three sections: two mass-separating quadrupole sections separated by a central quadrupole (or a higher multipole) section whose function is to contain the ions during fragmentation. Quadrupole mass spectrometers are capable of unit mass resolution and mass accuracy of 0.1–1 Da and excel at quantitative measurements. Owing to the presence of two independent quadrupole sections, the triple quadrupole can be programmed for a variety of different scan modes in addition to the isolation of a peptide followed by obtaining a mass spectrum of the fragments.

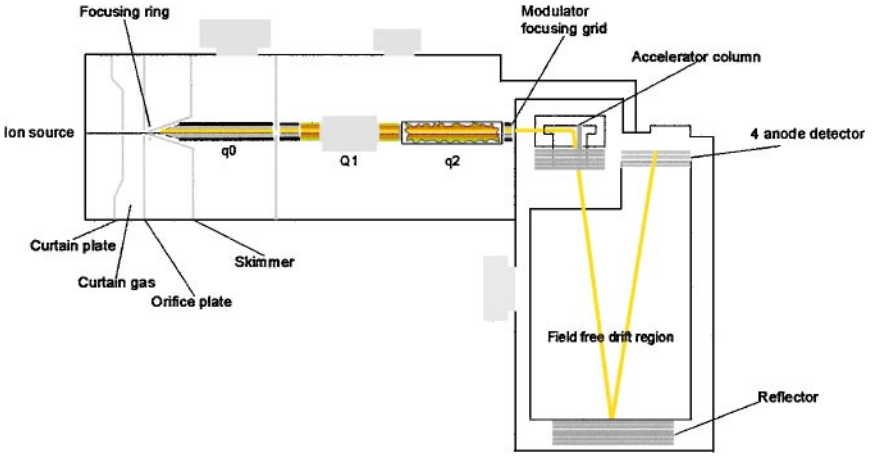


Figure 2 Schematic of a quadrupole TOF instrument. Ions are produced at atmosphere in the ion source shown to the left. After traversing a countercurrent gas stream (curtain gas), the ions enter the vacuum system and are focused into the first quadrupole section (q_0). They can be mass-separated in Q_1 and dissociated in q_2. Ions enter the time-of-flight analyzer through a grid and are pulsed into the reflector and onto the detector, where they are recorded. There are 14,000 pulsing events per second.

In recent years the triple quadrupole has begun to be complemented by the quadrupole time-of-flight instrument (Figure 2) in which the third quadrupole section is replaced by a TOF analyzer (12, 13). Its main advantages are that it provides the high accuracy for mass and high resolution typical of modern time-of-flight instruments, resulting in unambiguous determination of charge state and very high specificity in database searches (see below).

Ion-Trapping Instruments

Ions can also be trapped for analysis in three-dimensional electric fields. These ion traps capture the continuous beam of ions up to the limit of their space charge (14–16). This is the maximum number of ions that can be introduced into the instruments without distorting the applied field. The ions are then subjected to additional electric fields, which eject one ion species after another from the trap, and are detected, to produce a mass spectrum. For a MS/MS experiment, all except the single desired ion species are ejected first. Then the remaining ion species is fragmented and its products are analyzed. Several sequential steps of MS/MS can in principle be performed, allowing detailed fragmentation studies. Owing to the operating principle of the trap, the lower end of the fragment mass range (below about a quarter of the parent ion mass) cannot be observed. Ion traps are compact and versatile instruments whose operation has been highly automated. Even though in theory very high performance can be achieved, in practice mass

accuracy and resolution do not achieve the levels of the quadrupole time-of-flight instruments.

Another version of the trapping principle is embodied in the Fourier transform ion cyclotron MS (FTMS). Here the ions are captured in a combination of electric fields and a very strong magnetic field, necessitating a very large magnet. Although sensitivity and mass accuracy can be outstanding (17, 18), this complex instrument is not currently in widespread use in biological mass spectrometry.

Other Mass Spectrometers

In addition to the main types of mass spectrometers described above, other combinations are under development. For example, the MALDI–quadrupole TOF (MALDI-qTOF) instrument consists of a MALDI ion source coupled to the quadrupole TOF analyzer (19). The attraction of this instrument, if it can achieve sufficient sensitivity for demanding proteomic applications, is that it combines the ability of MALDI to analyze a large number of samples with the ability to sequence any selected peptides. However, MALDI produces far fewer ions than electrospray and they are singly charged, which leads to fragmentation patterns that are more complicated than those of the typically multiply charged ES ions.

MALDI can also be combined with a two-section TOF instrument separated by a fragmentation chamber, a so-called TOF-TOF arrangement (20). This instrument should have the advantage of very fast analysis times. As in the case of the MALDI quadrupole TOF, a challenge for this design is the low number of MALDI-produced ions. Additionally, it is more difficult to obtain high-resolution and highly accurate fragmentation spectra after an initial TOF separation compared to the well-defined ion beam produced by the MALDI–quadrupole TOF instrument.

PROTEIN PURIFICATION AND PREPARATION

The up-front isolation procedures can have the most significant impact on the outcome of an MS-based investigation. For example, sensitivity of the overall procedure is usually determined more by the purification strategy than by the sensitivity of the MS instrument per se. Typically, protein purification starts with a whole-cell lysate and ends with a gel-separated protein band or spot. MS analysis is usually carried out on peptides obtained after enzymatic degradation of these gel-separated proteins. In special cases, the intact proteins are analyzed or the gel electrophoretic step is omitted by digesting a collection of proteins in solution and analyzing the resulting complex mixture of peptides.

Special Considerations for Protein Preparation Methods

In principle, any of the classical separation methods such as centrifugation, column chromatography, and affinity-based procedures can precede the final gel electrophoresis. As long as the proteins of interest can be adequately resolved, it is

best to minimize the number of separations. Generally, silver-stained amounts are necessary for successful MS identification of proteins [5–50 ng or 0.1–1 pmol for a 50-kilodalton (kDa) protein], but even higher sensitivities have been achieved by specialized groups. It is important to minimize contamination with keratins, which are introduced by dust, chemicals, handling without gloves, etc, as the keratin peptides can easily dominate the spectrum. Most detergents and salts are incompatible with both 2-D gels and MS. Therefore, dialysis of the sample may be required. If the protein can be eluted from reversed-phase media, the best sample preparation is achieved on small, low-pressure traps that can be incorporated into MS injection ports. In affinity-based protocols it is important that the bait is pure, as contaminating proteins, for example bacterial proteins in a Glutathione S. transferase (GST) fusion preparation, will hinder analysis.

Digestion and Preparation of Gel-Separated Proteins for MS Analysis

For MS-based analysis, most of the detergents and salts are eliminated in the gel washing procedure. Nevertheless, the protein should be as concentrated as possible in the gel, to avoid excessive background in the MS analysis. Pooling of spots is not necessarily advantageous, as both protein and background will increase. Coomassie staining, silver staining, or radioactive labeling, which are all compatible with MS analysis, can visualize proteins. Cross-linkers and harsh oxidizing agents should be avoided as they interfere with extraction of peptides from the gel or may chemically modify the peptides (21). Silver staining provides adequate sensitivity, but it should be recognized that it has a narrow linear range.

Protein bands are excised from the gel and subjected to reduction, alkylation, several washing steps, and finally enzymatic digestion followed by peptide extraction. (For a recent protocol see 22.) A small portion of the resulting peptide mixture can be directly used for MALDI peptide mapping. For electrospray analysis, and often for MALDI analysis, peptides are desalted and concentrated. This can be performed by columns of reversed-phase material (50 nl to 1 μl in packing volume) in nanospray needles or gel loader tips or by injection into liquid chromatography MS columns (see below). The peptides are then eluted onto the MALDI target or into a nanoelectrospray spraying needle or are loaded onto a microcapillary column.

ANALYSIS OF INTACT PROTEINS

Several uses of mass spectrometry involve the characterization of recombinant proteins. For example, the glycosylation and disulfide bonding pattern of therapeutic proteins, such as growth factors, can be studied in detail. A mass spectrum of the intact protein provides the precise molecular weight of the major and minor forms of the protein, data that cannot always be gained from peptide mapping. However, larger proteins are typically heterogeneous, making molecular weight determination difficult.

Electrospray is the method of choice for determining molecular weight of proteins, as MALDI results in broad peaks and low sensitivity for proteins above about 30 kDa. As mentioned above, proteins need to be free of detergents and salts, which is usually accomplished by reversed-phase chromatography. Formic acid can be used to solubilize proteins for electrospray analysis. Figure 3 shows the mass spectrum of a small protein obtained on a high-resolution instrument (FTMS).

Identification of intact proteins from cell lysates by molecular weight alone is difficult for several reasons. Sensitivity of ESMS for large molecules is poorer than for peptide analysis because the signal is distributed over many charge states and the heterogeneity of the protein similarly splits the signal into many components. More fundamentally, the molecular weight of a protein cannot be predicted precisely from its database entry, because of N- and C-terminal processing, post-translational modifications, and chemical modifications introduced during sample purification (for example, oxidation of methionines). Therefore, even a precise molecular weight by itself will not allow identification of the protein. Nevertheless, in a recent study several hundred proteins were recorded in a single experiment in which isoelectric capillary electrophoresis of lysates of an *Escherichia coli* cell was combined with FTMS (18). Identification of the proteins was not achieved in this experiment for the above-mentioned reasons.

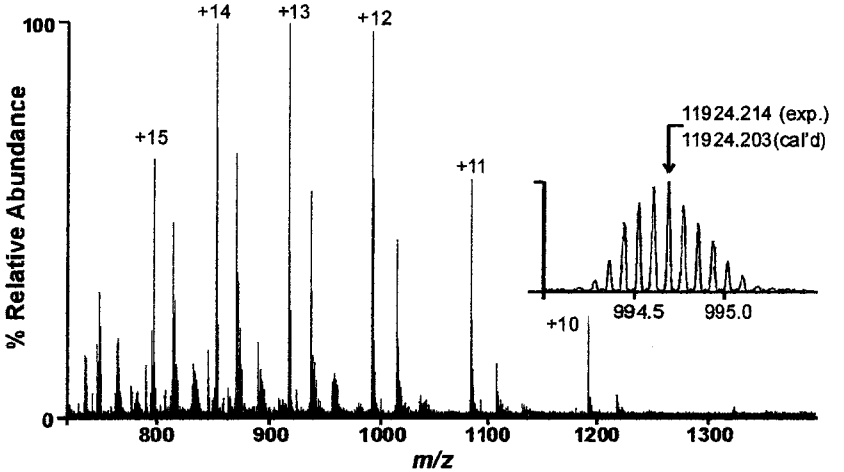


Figure 3 High-resolution Fourier transform ion cyclotron resonance mass spectrometry (FTMS) for accurate determination of mass of intact proteins. Alligator parvalbumin (APV-1) was isolated and solubilized in 50:50 methanol:water with 0.1% acetic acid to a final concentration of 100 fmol/μl. This solution was introduced into a 7-tesla FTMS instrument at a constant flow rate of 50 nl/min. High-resolution mass spectra were acquired using dipolar chirp excitation followed by heterodyne detection. The mass of the most abundant isotope for the 12+ charge state was determined to be 11,924.06 Da (external calibration). This agreed to within 12 ppm with the true mass based on the known amino acid sequence. (Figure courtesy of Jarrod Marto and Susan Martin.)

PEPTIDE SEQUENCING BY TANDEM MASS SPECTROMETRY

The sequence of peptides can be determined by interpreting the data resulting from fragmenting the peptides in tandem mass spectrometers (23, 24). In this technique, one peptide species out of a mixture is selected in the first mass spectrometer and is then dissociated by collision with an inert gas, such as argon or nitrogen. The resulting fragments are separated in the second part of the tandem mass spectrometer, producing the tandem mass spectrum, or MS/MS spectrum. In the instruments in use today, multiple collisions impart energy onto the molecule until it fragments. (This is low-energy fragmentation, in which any single hit is not sufficient to break the peptide bond. In high-energy fragmentation, the molecules have higher velocity and a single hit can break bonds.)

As shown in Figure 4, several bonds along the backbone can be broken by the collisions. The most common ion types are the b and the y ions, which denote fragmentation at the amide bond with charge retention on the N or C terminus, respectively. Most proteomics experiments are performed with tryptic peptides, which have arginyl or lysyl residues as their C-terminal residues. In this case, y ions are the predominant ion series observed.

Most peptide sequencing is performed on electrosprayed ions. These ions generally have a charge state corresponding to the number of positively charged amino acids plus the charge formally localized at the N terminus of the peptide. Thus tryptic peptides are often doubly charged, or triply charged if they contain a histidyl residue. Larger peptides are multiply charged, and their fragmentation spectra often contain multiply charged ion series as well.

Tandem mass spectra are usually interpreted with computer assistance, or matched against databases directly (see below). In very high quality spectra it

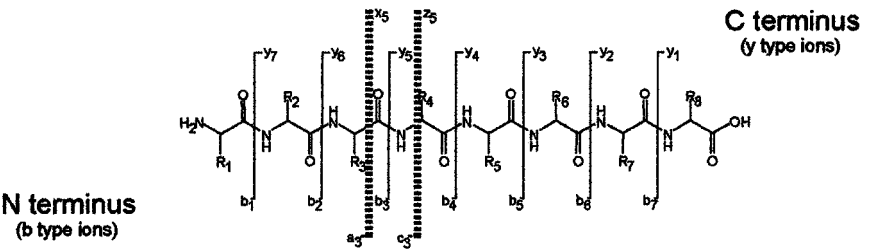


Figure 4 Nomenclature for the product ions generated in the fragmentation of peptide molecules by tandem mass spectrometry. Collision-activated dissociation (CAD) causes a single cleavage to occur more or less randomly at the various amide bonds in the collection of peptide molecules. This process generates a series of fragments that differ by a single amino acid residue. Ions of type y contain the C terminus plus one or more additional residues. Ions of type b contain the N terminus plus one or more additional residues. Additional ion types correspond to cleavages at different positions in the backbone (dashed lines).

is possible to interpret the fragmentation ladders (the b and the y ion series) from the low mass end through to the highest mass ion. For example, the y ion series of tryptic peptides will start with masses $y_1 = 147$ (Lys) or 175 (Arg). The next fragmentation peak in the y ion series, the y_2 ion, differs by the mass of an amino acid residue and thus "spells out" the next amino acid. Similarly, the b ion series starts with b_1 for the N-terminal amino acid and is traced upward in molecular weight. Note that the b and y ions are not distinguishable a priori. Ideally, a complete set of b and y ions will (doubly) confirm the entire peptide sequence. In practice, not all fragment ions are present at detectable levels and fragments also arise by double fragmentation of the backbone (internal fragment ions). Therefore, it is often possible to interpret part but not all of the sequence with confidence.

Other features of the MS/MS spectrum include immonium ions, which arise by double cleavage of the peptide backbone, N-terminal and C-terminal to the amino acid residue. These immonium ions can indicate or confirm the presence of individual amino acids.

Fragmentation is not equally likely along the entire length of the peptide backbone. For example, fragmentation between the two N-terminal amino acids is energetically unfavorable and therefore the b_1 ion is often not observed. The b_2 ion, however, and its companion, the a_2 ion, at a 28-Da (CO) mass difference, is usually very prominent. Likewise, fragmentation at the C-terminal bond of a proline residue is weak but cleavage at the N-terminal side is usually very prominent.

Another feature that can frequently be observed is the switchover from a C-terminal to a N-terminal series because of charge retention of a positively charged amino acid residue, such as histidine.

Bond breakage of the doubly charged tryptic peptides most often used in protein characterization is believed to proceed via the localized charge on the arginines or lysine residue on the C terminus and the delocalized proton formally located on the N terminus, which induces the amide bond breakage along the backbone. The arginyl residue strongly localizes the proton, and peptides containing internal arginines often result in MS/MS spectra that are very difficult to interpret.

As mentioned above, fragmentation of large, multiply charged ions often leads to multiply charged ion series. High resolution is advantageous in studying these ions because it allows direct assignment of the charge state of fragments based on the spacing of the carbon isotope peaks (for example, they are spaced 1/3 of a dalton apart if they are triply charged). Larger peptides often fragment efficiently and provide long ion series, but because the precursor ion intensity is distributed over several charge states, sensitivity may not be as high.

Mass spectrometers in use today cannot distinguish between isoleucine and leucine, which have the same mass (though the distinction can in principle be made by using the different retention time of leucine- and isoleucine-containing peptides during chromatography or the side chain fragmentation in high-energy collisions). The glutamyl and lysyl residues have the same nominal mass but can be distinguished by their mass difference of 0.036 Da on modern TOF or FTMS instruments.

PROTEIN IDENTIFICATION BY DATABASE SEARCHING

A key advance in biological mass spectrometry was the development of algorithms for the identification of proteins by mass spectrometric data matched to a database, originally using a set of peptide masses and now increasingly using the fragmentation spectra of the individual peptides. For the reasons described in previous sections, obtaining the complete sequence of a peptide from the tandem mass spectrum was time consuming at best and often impossible. With the availability of the complete sequence of an increasing number of model species, the peptide sequencing problem, formerly a holy grail in biological mass spectrometry, is reduced to a database correlation, enabling automation and the scaling up of proteomics experiments.

Peptide Mass Fingerprinting

In this method, a "mass fingerprint" is obtained of a protein enzymatically degraded with a sequence-specific protease such as trypsin. This set of masses, typically obtained by MALDI-TOF, is then compared to the theoretically expected tryptic peptide masses for each entry in the database. The proteins can be ranked according to the number of peptide matches (see Figure 5). More sophisticated scoring algorithms take the mass accuracy and the percentage of the protein sequence covered into account and attempt to calculate a level of confidence for the match (25–27). Other factors can also be included, such as the fact that larger peptides are less frequent in the database and should therefore count more when matched. The accuracy obtained in the measurement of peptide mass strongly influences the specificity of the search (7, 28). When high mass accuracy (10 to 50 ppm) is achieved, as a rule at least five peptide masses need to be matched to the protein and 15% of the protein sequence needs to be covered for an unambiguous identification. After a match has been found, a second-pass search is performed to correlate remaining peptides with the database sequence of the match, taking into account possible modifications.

Mass fingerprinting can also resolve simple protein mixtures, consisting of several proteins within a roughly comparable amount. For example, databases can be searched iteratively by removing the peptides associated with an unambiguous match (29).

Generally, peptide mass fingerprinting is used for the rapid identification of a single protein component. Protein sequences need to be in the database in substantially full length. Isoforms can be differentiated from each other, if peptides covering the sequence differences appear in the peptide map. In our experience, proteins from organisms with fully sequenced genomes can be identified with a 50–90% success rate when at least a few hundred femtomoles of gel-separated protein are present.

Searching with Tandem Mass Spectrometric Data

Databases can also be searched by tandem mass spectrometric data obtained on peptides from the proteins of interest. Because the tandem mass spectra contain

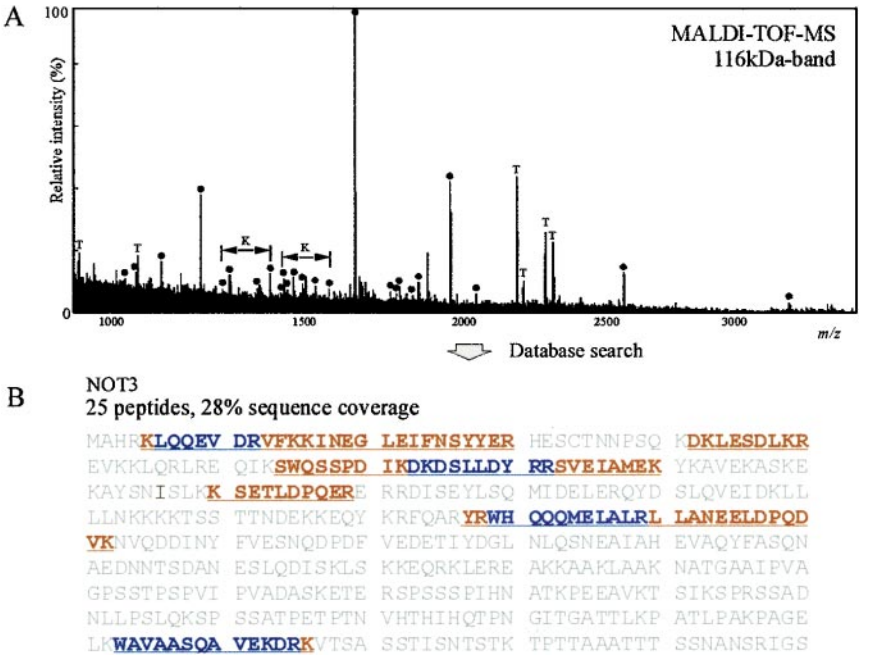


Figure 5 MALDI peptide map and identification of a protein. A 116-kDa band was excised and subjected to tryptic digestion in gel. A few percent of the supernatant after digestion was applied to a thin film matrix on a metal target. (*A*) The resulting peptide mass spectrum. Peaks resulting from trypsin autodigestion are marked, as are peaks spaced exactly by the mass of the lysyl or arginyl residues, denoting peptides ending in a double tryptic cleavage site. (*B*) A graphical representation of the database hit. Matched peptides appear in red, and sequences that were covered by two peptides in blue.

structural information related to the sequence of the peptide, rather than only its mass, these searches are generally more specific and discriminating.

Several approaches exist. The peptide sequence tag method (30) makes use of the fact that nearly every tandem mass spectrum contains at least a short run of fragment ions that unambiguously specifies a short amino acid sequence (see Figure 6). As few as two amino acids can be combined with the start mass and the end mass of the series, which specify the exact location of the sequence in the peptide and the known cleavage specificity of the enzyme. Such a peptide sequence tag will then retrieve from the database one or a few sequences whose theoretical fragmentation pattern is matched against the experimental one. The procedure can be automated and is highly specific, especially when performed using instruments with a high accuracy for mass, such as the quadrupole TOF instrument.

Other methods do not attempt to extract any sequence information at all from the MS/MS spectrum (31). Instead, the experimental spectrum is matched against a calculated spectrum for all peptides in the database. A score is given to determine

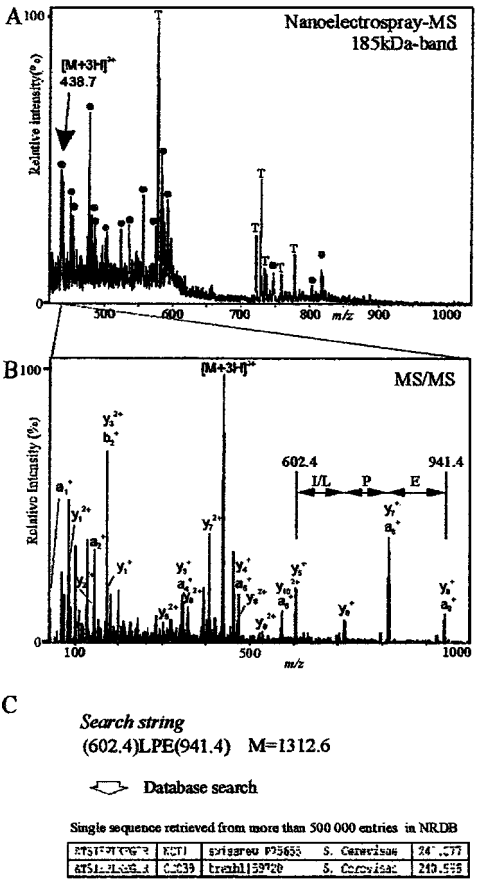


Figure 6 Identification by peptide sequence tag. (*A*) The nanoelectrospray mass spectrum of the tryptic, micropurified peptides derived by in gel digestion of a protein band. Tryptic autolysis products are marked. Peptides sequenced are marked by a filled circle. (*B*) The result of fragmenting the triply charged ion at $m/z = 428.7$. Three amino acids, isoleucine/leucine, proline, and glutamic acid, are apparent in the higher mass part of the spectrum. (*C*) Once the start mass, the sequence, and the end mass, together with the parent mass, are converted into a peptide sequence tag and searched in a nonredundant database, two identical peptide sequences are retrieved. The calculated fragmentation spectrum for this sequence annotates the experimental spectrum, verifying the peptide match. Further peptides apparent in panel *A* were selected and independently verify the protein match. Finally, peptides that differ between the two protein database entries are calculated and the corresponding *m/z* values are isolated and fragmented, allowing differentiation between the two isoforms (intelligent data acquisition).

how much the tandem mass spectrum agrees with the calculated sequence. Another score indicates how differently the next most similar sequence in the database fits the spectrum. Although this method can be highly automated, the sequences need to be verified by manual inspection unless the score is very high. In a typical liquid chromatography MS/MS experiment on an ion trap, about 10–20% of MS/MS spectra may require no further interpretation, whereas a much larger percentage can be verified manually.

Tandem mass spectrometry allows direct analysis of protein mixtures. Bands from one-dimensional gels, when analyzed with the high sensitivity of the mass spectrometer, often turn out to contain several proteins. In such a case, the software reports a list of proteins, each matched by one or several peptides. In extreme cases, crude protein mixtures can be reduced to peptides, and the peptides fragmented and searched in databases (32). In this way, large numbers of proteins, up to hundreds, can be identified all at once (see LC-MS/MS below).

Searching Protein, Expressed Sequence Tag, and Genome Databases

Three types of sequence databases are searched by mass spectrometric data. Nonredundant protein databases contain the known set of full-length protein sequences, extracted from the major sequence repositories and purged of duplicates. For example, nrdb, maintained at the European Bioinformatics Institute, contains more than 500,000 sequences as of this writing. These databases can be searched both by mass fingerprint and tandem mass spectrometric data. Matches can quickly be followed up via links to annotated protein databases such as SwissProt or in-depth curated ones such as the Yeast Protein Database (33). Expressed sequence tag (EST) databases, such as dbEST at the National Center for Biotechnology Information, contain millions of short one-pass sequences from random sequencing of cDNA libraries. These can be searched with appropriate software, usually by translating into the six reading frames. Even though EST databases are highly error prone and cover only a part of the gene, it is our experience that virtually all proteins encountered in proteomics projects can be correlated to their respective EST entries based on minimal mass spectrometric information (34, 35). It is significantly more difficult to follow up on a list of EST matches obtained in a proteomics experiment because the EST entries need to be assembled and searched for homology. Finally, genome databases can also be searched with mass spectrometric data. Surprisingly, it has been found that peptides can be matched in the raw genome sequence, without any assumptions about the reading frame or likely coding regions and without translation into amino acid sequence (Küster, Andersen, Mortensen & Mann, unpublished information). The advantages of searching databases of completely sequenced genome directly are that each peptide must be present by definition, that genome sequences are of very high quality, and that often the mass spectrometric data can help to define the structure of the gene, such as start and stop and intron-exon structure.

Database Searching and Protein Modifications

Protein modifications do not present an obstacle to identification. Of a typical 50 peptides (for a 50-kDa protein) generated by tryptic digestion, only a few will be modified. As described above, only a small number of peptides are required for unique matching to a database entry, especially in the case of data from tandem mass spectrometry; therefore even extensive modification only marginally increases the difficulty of protein identification.

Modifications may also be discovered while searching databases. For example, a phosphopeptide can be correlated to a peptide sequence in the database with an additional mass increment due to the phosphogroup (80 Da). In the algorithm for the peptide sequence tag, a mass difference at either side of the tag sequence can be allowed. For example, the tag sequence and the mass to the C terminus of the peptide could agree with the database entry. However, the mass to the N terminus could be larger by 80 Da. In this case there would have to be a modification yielding a mass difference of 80 Da between the N terminus of the peptide and the start of the tag sequence.

The peptide fragmentation spectra can also be calculated for all possible combinations of common modifications. For example, serines in suspected phosphopeptides can be substituted in turn with phosphoserines when calculating fragmentation spectra. Because of the increase in possible peptide fragmentation spectra to consider and the increase in computation time, this search is usually not performed for the whole database but only for a small set of sequences, including the protein sequence already identified in the database from unmodified peptides.

LIQUID CHROMATOGRAPHY AND TANDEM MASS SPECTROMETRY

Principles

Liquid chromatography (LC) coupled to tandem mass spectrometry, called LC-MS/MS (sometimes abbreviated simply as LC-MS), is a powerful technique for the analysis of peptides and proteins. This methodology combines efficient separations of biological materials and sensitive identification of the individual components by mass spectrometry. Complicated mixtures containing hundreds of proteins can be analyzed directly even when concentration levels of different proteins vary by orders of magnitude. LC-MS/MS can be used alone or in combination with 1-D or 2-D electrophoresis, immunoprecipitation, or other protein purification techniques.

Although numerous methods for coupling liquid chromatography to mass spectrometry have been explored (reviewed in 36), it is electrospray ionization that has transformed LC-MS/MS into a routine laboratory procedure sensitive enough to analyze peptides and proteins at levels interesting in biological research. As described above, ES requires a continuous flow of liquid, and signal strength is

independent of flow rate. In order to obtain maximum sensitivity, research efforts have focused on coupling nano-scale LC at submicroliter flow rates to the highly sensitive micro-scale ES interface (37–39). (We refer to this technique as nanoLC-MS/MS for its nanoliter per minute flow rate, but it is also called microES, microcapillary LC, or picospray.) Currently, detection limits of a few femtomoles of peptide material loaded on the column make this technique compatible with silver-stained, fluorescently labeled, or faintly stained Coomassie gel bands and capable of detecting proteins and peptides present at a low copy number per cell.

In a typical LC-MS/MS experiment, the analyte is eluted from a reversed-phase column to separate the peptides by hydrophobicity, and is ionized and transferred with high efficiency into the mass spectrometer for analysis. A large amount of data regarding individual species in a complicated mixture is generated. For example, the peptide ligands associated with the human major histocompatibility complex (MHC) class I molecules, HLA-A2.1, are a mixture of approximately 10,000 different peptide species (40). As shown in Figure 7, an aliquot representing the amount isolated from 1×10^8 cells was loaded onto a nanoLC column and eluted into an ion trap mass spectrometer using a long gradient. Mass spectra were acquired over the mass range 300–2000. The data acquired contain molecular weight information on the peptide species and their amounts. The ion current for each scan can be summed and plotted as a function of time. This display is a total ion current chromatogram (TIC), shown in Figure 7*A*, and is similar to a UV chromatogram. Postacquisition, the data can be interrogated to reveal the ion current recorded at a particular *m/z* (molecular weight), or a selected ion current chromatogram (SIC) (Figure 7*B*). Any individual peptide can be sequenced without further purification by isolating the eluting peptide, fragmenting it, and obtaining the MS/MS spectrum (Figure 7*D*). In this manner a large number of peptides can be sequenced in a single LC-MS/MS run, even those that have the same molecular weight if they differ in hydrophobicity. In practice, the mass spectrometer is often programmed to perform one scan to determine the peptide masses and then to sequence the three to eight most abundant peptides (data-dependent acquisition). The separation principle is almost always reversed-phase high-performance liquid chromatography (HPLC) as applied elsewhere in protein chemistry; the only difference is that the dimensions and flow rates are much smaller (see below). Peptides elute with a typical peak width of 30 seconds. Higher separation efficiencies could in principle be achieved with coated open tubular capillaries, 1–5 μm internal diameter, which have been used for high-sensitivity analysis, but the small amount of wall surface area results in a lower capacity factor. Capillary zone electrophoresis (CZE) is another technique that has been coupled to electrospray. Although CZE-MS achieves some of the same advantages as nanoLC-MS/MS, in practice it is less rugged and is not in general use.

Instrumental and Practical Considerations

Low-flow-rate LC (<500 nl/min), or nanoLC, once thought to be a specialized and difficult technique, has become a routine procedure in many laboratories over the

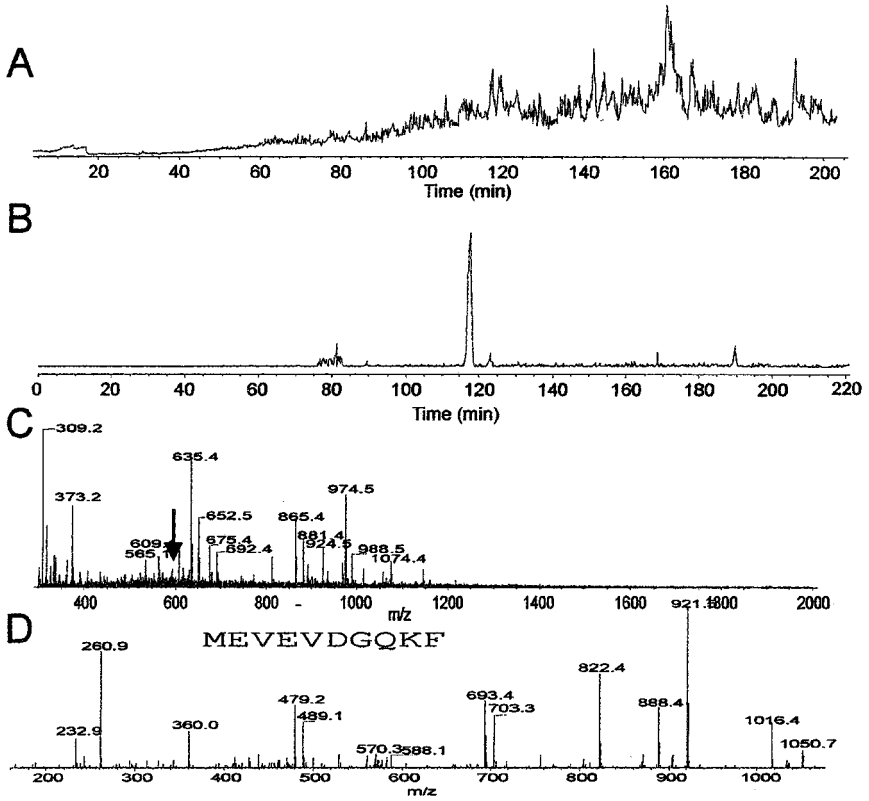


Figure 7 LC-MS analysis of a complex mixture of peptides isolated from human MHC class I molecules. An aliquot of peptides was loaded onto a nanoLC column and eluted with a gradient of acetonitrile and 0.1 M acetic acid. Spectra over the mass range 300 to 2000 were acquired every 1.5 seconds. (*A*) The sum of all the ion current recorded at the detector plotted as a function of scan number, a total ion current chromatogram (TIC). (*B*) The data replotted to show the ion current for $m/z = 591.5$, called a selected ion current chromatogram (SIC). (*C*) Approximately 20 mass spectra at elution time 117 min were summed to generate a single mass spectrum, which shows all the peptides that co-elute with the peptide at $m/z = 591.5\,(\mathrm{M}+2\mathrm{H})^{2+}$ [doubly charged peptide indicated by $(\mathrm{M}+2\mathrm{H})^{2+}$]. (D) The MS/MS spectrum for the $(\mathrm{M}+2\mathrm{H})^{2+}$ ion at m/z 591.5 and the interpreted amino acid sequence. (Figure courtesy of Jarrod Marto and Susan Martin.)

last few years. The low flow rates required are achieved by modifying a conventional LC system, transfer lines, gradient mixture, and use of a precolumn splitting device (41) or by the use of specially designed nanoLC systems. An online UV detector is incorporated into the system, if desired. Typical nanoLC columns are 50–100 μm in internal diameter and are packed with polymeric or silica-based, C18 coated, stationary phases with typical particle sizes in the 3- to 10-μm range. The smaller the column diameter, the lower the flow rate for the same chromatographic

separation and hence the higher the sensitivity. Column diameters of 75 μm are the current compromise between ultimate sensitivity and trouble-free operation on a routine basis. Columns are available from commercial vendors or are packed in individual laboratories. As in conventional LC, the sample can be introduced onto the column via loop injections of the sample. Auto samplers are available that inject submicroliter volumes. To minimize losses associated with handling of the sample, an alternative method is to displace the sample directly from the sample tube onto the column by pneumatic displacement (bomb loading). Ideally, the sample is loaded at a higher flow rate (microliters per minute) onto a trap column and is then eluted onto the separation column. This arrangement has the further advantage of leading to fewer problems of plugging in these very fine columns.

An electrical connection needs to be made between the liquid and a power supply in order to supply the charges to the electrospray process. In one type of connection the electrospray needle, or emitter, is coated with a conductive material and the voltage is applied directly. Alternatively, a liquid junction, approximately 3 nl, is purposefully formed in a metal union (stainless steel, gold, titanium) between the exit of the LC column and the electrospray needle (42) or by applying the spraying voltage prior to the column (43). A typical nanoLC interface is shown in Figure 8. For high-sensitivity LC-MS/MS applications, careful consideration must be given to solvent purity. For example, even though UV-detectable trace contaminants may not be present, the solvents may contain ionizable impurities that reduce the final signal to noise of the analysis. The nanoLC system couples to various mass spectrometers such as triple quadrupole, quadrupole TOF, ion traps, and FTMS instruments.

What is the typical performance one can achieve? For nanoLC systems, sensitivity at 1–10 fmol in LC-MS/MS mode is routine on a triple quadrupole mass spectrometer (40) and has been reported in the low attomole range for selected ion monitoring. The newer quadrupole TOF instrument promises to provide a large improvement in sensitivity in MS/MS mode over the triple quadrupole instrument, in addition to increased mass resolution. The ion trap mass spectrometer has sensitivity in the full-scan MS mode of 1–5 fmol and is mainly limited by chemical noise introduced into the trap. However, in MS/MS mode, increased duty cycle of the ion trap instrument results in an improvement of ultimate sensitivity reported in the 10- to 50-attomole range (44). This feature of the ion trap can be used only when sequencing a particular, known mass during an LC-MS/MS run (for example, when sequencing a putative phosphopeptide or when the mass has been already been determined by a more sensitive full-scan method of mass spectrometry such as FTMS). Recently, automated variable-flow LC, 5–200 nl/min (45), also known as peak parking (46), has been employed to improve sensitivity to the 10- to 50-attomole level in MS/MS experiments on an ion trap mass spectrometer. On the FTMS, LC-MS sensitivities at the 10-attomole level with a dynamic range of 10^3 have been reported. Sample carryover does not appear to be a problem in the analysis when high-sensitivity applications are dedicated to a single instrumental setup. It should be noted that all these sensitivity numbers relate to material applied to the column rather than protein material in a gel band.

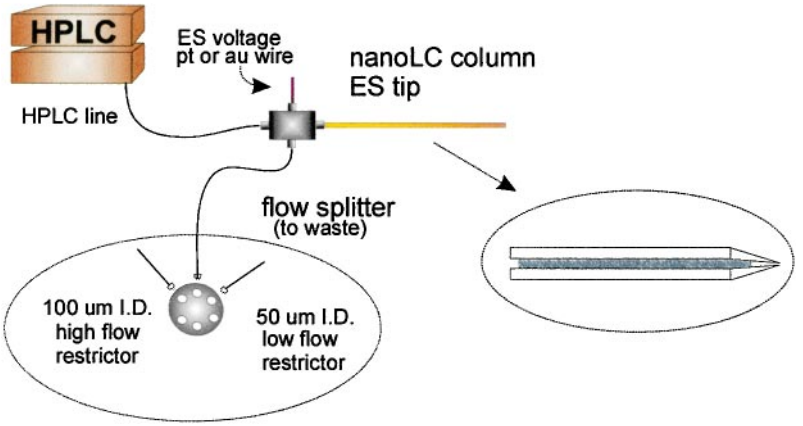


Figure 8 Schematic diagram of nanoLC ES interface for mass spectrometry. In this setup, an HPLC gradient flowing at 200 μl/min is split 1:1000 using a micro-cross precolumn splitting device to achieve an operating flow of 200 nl/min. The split ratio is varied by adjusting the length and/or internal diameter (I.D.) of the restrictive capillary on the flow splitter. Electrical contact for electrospray ionization (ESI) is made via a metal-liquid junction by inserting a gold or platinum wire into one arm of the micro-cross fitting. Typical voltages are 0.7–1.9 kV. The nanoLC column is inserted into the remaining arm of the micro-cross. (*Right*) Schematic of a fritless 360 μm outer diameter $\times$ 75 μm inner diameter LC column prepared by using a laser puller to generate a tip with a tapered internal diameter, approximately 5–10 μm, to retain the packing material. (*Lower left*) Schematic of an automatable apparatus for variable-flow LC; the flow can be directed to different sizes of restrictor capillaries.

Recent Advances in Instrumentation and Strategy

Recent improvements in instrument sensitivity, software control, automation, and data analysis tools provide unique high-throughput capabilities for protein analysis when using LC-MS/MS. One approach is based on the proteolytic digestion of mixtures of proteins and LC-MS/MS analysis of the peptides generated. Because peptides are now scrambled or multiplexed, it is no longer possible to identify proteins based on peptide mass profiling; instead, tandem mass spectral data must be generated and interpreted. Software control and automation accelerate the process of acquiring the MS/MS data. For example, in a typical data-dependent acquisition, the mass spectrometer generates a full-scan mass spectra to determine the molecular weights of the peptide species present and then acquires MS/MS spectra on three to five peptides. This cycle takes less than 15 seconds to complete. In this manner hundreds of MS/MS spectra can be generated in a single run. The data analysis tools (described earlier) are essential in these high-capacity experiments. The analysis of a complex mixture is shown in Figure 7 and Figure 13.

Ion exchange chromatography can precede HPLC to reduce the complexity of the peptide mixture (2-D chromatography), as in the direct analysis of the yeast

ribosome complex identifying 80 proteins in a single experiment (47, 48). In a similar experiment using off-line ion exchange chromatography and long-gradient nanoLC, Gygi et al (personal communication) identified more than 500 yeast proteins from yeast extract using high stringency for the database identifications. When applying these techniques to a single protein, high coverage of sequence can be achieved; for example, the variations in amino acid sequence that result from single mutations in hemoglobin can be analyzed (49). Membrane proteins and lipid molecules can also be analyzed by these techniques (50, 51).

ANALYSIS OF POSTTRANSLATIONAL MODIFICATIONS

Genomic methodologies provide significant information about gene structure as well as other events such as splicing. However, the vast array of posttranslational modifications commonly observed in proteins cannot be studied or be predicted accurately. Proteomic techniques are the only solution to definitively study post-translational modifications. We mainly restrict our discussion to protein phosphorylation and the methods in use to facilitate its detection by mass spectrometry.

Protein phosphorylation is a major conduit of information within the cell. It has been intensively studied over the last decade since kinases regulate most of the signals originating at the cell surface by phosphorylation. Serine and threonine phosphorylation of proteins is the more common form, occurring in about 99% of cases, whereas tyrosine phosphorylation occurs in less than 1% of cases (52). This is also reflected in the genomic organization of several organisms in which serine/threonine kinases far outnumber tyrosine kinases (53, 54). In spite of the fact that a large number of kinases and phosphatases have been identified, very few physiological substrates have been described. The substrates known to date were mainly discovered using strategies that rely on in vitro phosphorylation activities of the kinase in question. Prediction of putative phosphorylation sites from so-called consensus sites generally yields a large number of false positives, making it necessary to confirm phosphorylation of residues in vivo.

At least in theory, owing to its sensitivity of detection, mass spectrometry can be the tool of choice in identification of phosphorylation sites on proteins in vivo. However, several issues still remain to be resolved. The first is that the protein of interest must be enriched sufficiently to provide enough material for analysis. This is mainly due to the fact that whereas identification of the protein in question by database searching requires sequencing of only one or two peptides without any further sample manipulation, determining phosphorylation sites may require more time and additional preparative steps (see below). Second, since tryptic peptides are usually analyzed in the mass spectrometer, the phosphorylated peptides may need to be separated from nonphosphorylated peptides, the presence of which may obscure the signal from the phosphorylated peptides. Last, since determination of the exact phosphorylated residue may still be difficult, it might be worthwhile to identify the relevant peptides that are phosphorylated by treatment with phosphatases and observe the mass difference of the peptides before and after treatment.

Numerous studies have reported identification of phosphorylation sites derived from model or gel-isolated proteins (55–61). The following methods are in use today for facilitating specific detection of phosphopeptides—their use is not mutually exclusive (see Figure 9).

1. Enrichment of phosphopeptides by immobilized metal affinity chromatography (IMAC): Gallium- or iron-chelated affinity columns can be used to preferentially bind to the negatively charged phosphate groups (61, 62). The disadvantage is that they may bind to nonphosphorylated residues such as glutamic and aspartic acid that also carry a negative charge. We have recently found that these columns can be used to enrich for some phosphorylated proteins as well (M Mann, M Grønborg, ON Jensen & A Pandey, unpublished data).
2. Treatment with alkaline phosphatase: Complex phosphopeptide mixtures can be extracted from the MALDI target where they are mixed with the matrix, digested with alkaline phosphatase, and reanalyzed by MALDI (63). A decrease in the mass of a peptide by 80 Da implies the presence of one phosphate group. If only one or two proteins are present in the sample, the peptide mass after phosphatase treatment can be used to predict which peptide was phosphorylated.
3. Mass spectrometric methods for sequencing phosphopeptides:
   A. Phosphopeptides can be sequenced using tandem mass spectrometers in the positive mode in the same way as other peptides. The fragmentation spectrum will reveal a loss of phosphoric acid (H_3PO_4) due to a β-elimination reaction in the case of phosphoserine- and phosphothreonine-containing peptides (98-Da loss for singly charged species) whereas phosphotyrosine residues, which are less susceptible to β-elimination, are generally more stable. The location of phosphoserines can thus be identified either by a mass difference of two successive fragments of (87 + 80) Da, i.e. 167 Da, or as 69 Da due to its β-elimination product, dehydroalanine. Similarly, the location of phosphothreonines can be identified as a mass difference of (101 + 80) Da or as 83 Da due to its β-elimination product dehydroamino-2-butyric acid. Phosphotyrosines can be localized by a mass difference of (163 + 80) Da or 243 Da. The fragmentation behavior of these phosphopeptides depends somewhat on the tandem mass spectrometer used.
   B. Precursor ion scanning in the negative mode to detect phosphopeptides under neutral or alkaline conditions: In this method, the second mass spectrometer in a tandem MS is fixed at the mass of the reporter ion for the phospho group, $m/z = 79$. Scanning through the mass range with the first mass spectrometer will result in a signal only if the peptide being fragmented yields the reporter ion. Once the phosphopeptide is identified, the remainder of the sample is—after reconstitution in acidic solution—analyzed in the conventional positive mode as in part A to

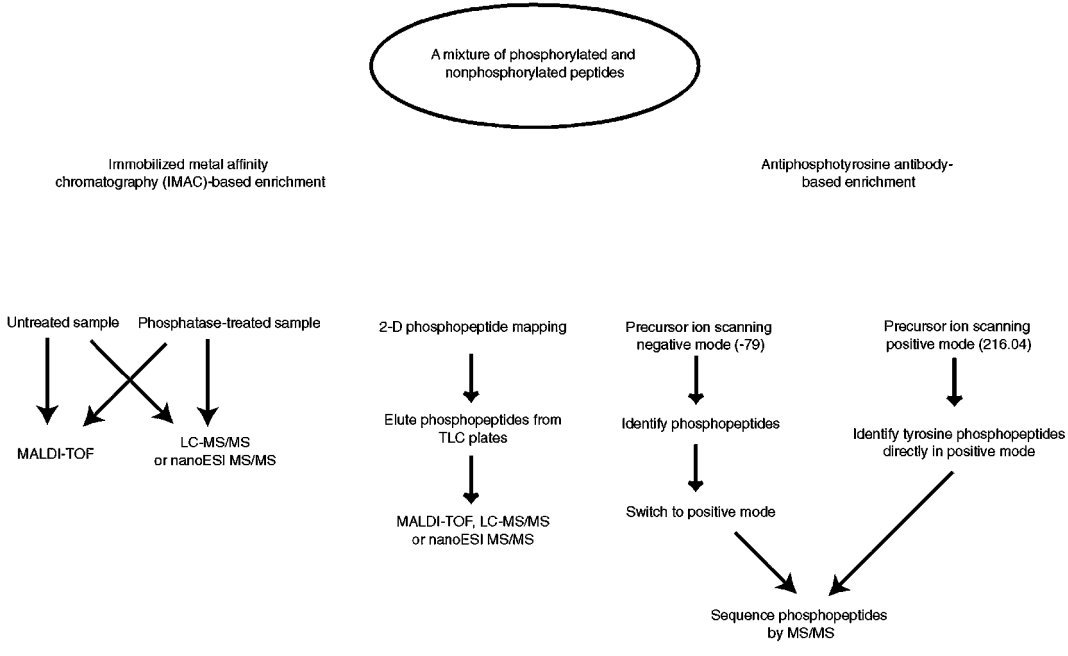


Figure 9 Phosphorylation analysis by mass spectrometry. The diagram shows various methods that can be employed to identify phosphorylation sites from tryptic digests of gel-isolated proteins. Phosphopeptides from the crude mixture can first be enriched by IMAC or antibody immunoprecipitation as shown or be analyzed directly. Many phosphopeptides can also be retrieved from the LC-MS and LC-MS/MS data, either by searching a database of MS/MS spectra and allowing for the presence of a phosphogroup on any serine, threonine, or tyrosine residues, or by filtering out ions that show loss of a fragment corresponding to the mass of the phosphogroup. ESI, electrospray ionization.

sequence the relevant phosphopeptide in order to identify the residue that is phosphorylated.

C. Parent ion scanning for phosphotyrosines—positive mode: Using the immonium ion of phosphotyrosine as a characteristic product ion, parent ion scanning can be applied in the positive ion mode (63a). This new scanning mode is possible because of the greater resolving power of newer mass spectrometers that can distinguish the characteristic reporter ion, the phosphotyrosine immonium ion at $m/z = 216.04$ from the y, b, and a series of amino acid doublets and triplets [such as b_2 series of asparagine threonine (NT) or glutamine serine (QS) doublets whose m/z is 216.09]. The main advantage of this method is that the parent peptide that is tyrosine phosphorylated can be identified and sequenced in the same positive mode. Unfortunately, the characteristic reporter ions for phosphoserine and phosphothreonine, which are much more labile, are not formed in positive mode, so this method cannot be used for these cases.

Other modifications most commonly found in proteins during routine proteomic analyses are acetylation, glycosylation, and products of N- or C-terminal processing by peptidases. Acetylated peptides are quite useful because they help in determining the amino terminus of mature proteins that are formed after posttranslational processing. Generally, all modifications that lead to a mass change can be analyzed by MS. Some modifications are very labile and care has to be taken that they are not reversed before analysis (for example, phosphohistidine is not stable in the acidic pH normally used to run electrospray). A specialized literature exists on several classes of modifications, such as disulfide bonds, glycosylation, xenobiotic modifications, and the like. The reader is referred to the biannual review of mass spectrometry in *Analytical Chemistry* (64).

QUANTITATIVE MASS SPECTROMETRY

As discussed above, the intensity of the mass spectrometric signal cannot easily be correlated with the amount of analyte present in the sample. However, in many proteomics problems, knowledge of quantitative changes is important. These changes usually have to be obtained from the staining intensities of proteins on gels, a labor-intensive method that is prone to error. Recently, stable isotope methods from small-molecule mass spectrometry have been introduced into MS-based proteomics and now allow relative changes to be determined. The principle of these methods is the incorporation of a stable isotope derivative in one of the two states to be compared. Stable isotope incorporation shifts the mass of the peptides by a predictable amount. The ratio of analyte between the two states can then be determined accurately by the measured peak ratio between the underivatized and the derivatized sample.

In a first application of this principle in proteomics studies, microbes were grown either on normal or ^{15}N media. The samples were then mixed and peak

ratios determined in MALDI analysis of gel-separated spots. Note that the sensitivity requirements for these approaches are somewhat higher than for straightforward identification because the peaks need to be well defined to allow isotope ratio measurements. The same principle has been applied in a variety of ways, including measurement of the relative amount of phosphorylation using stable isotope labeling in whole cells (65).

In an interesting twist on this method, the isotope label is attached during the chemical processing of the proteins, at the stage of blocking of the reactive cysteine groups (48). Instead of the normally employed iodoacetamide, a blocking group consisting of a biotin label and a linker containing either normal hydrogen or deuterium atoms was employed. Only cysteine-containing peptides are retrieved by biotin-avidin interaction chromatography, and the ratio of normal to deuterated peptides can be used to accurately determine the quantitative change of the peptide in the sample. Advantages of the method include the fact that the label is incorporated during protein workup and that the protein mixture is simplified because only cysteine-containing peptides are present. In contrast to two-dimensional gel electrophoresis, quantitative changes in proteins in very low abundance can be determined, provided sufficient material is purified. Limitations are that some proteins do not contain cysteines or do not contain cysteines in tryptic peptides in the preferred mass range for ESMS sequencing. To avoid this limitation, other methods to quantify differences in peptides employ a tag at the N terminus of peptide (66).

PROTEOMIC STRATEGIES AND EXAMPLES

Differential 2-D Gels Followed by Mass Spectrometric Protein Identification

Classical proteomic approaches have relied upon separating whole-cell lysates by 2-D gel electrophoresis. In fact, 2-D gel electrophoresis was almost synonymous with proteomics until quite recently. Except for simpler organisms such as *E. coli*, the hope that a combination of molecular weight and isoelectric point (pI) of the protein will be sufficient for identification of the protein has not been realized. More popular are scenarios termed expression proteomics in which two states such as cancerous versus normal tissue may be compared by 2-D gel electrophoresis (67). Mass spectrometric identification, although quite sensitive, is limited by the ability to reproducibly observe distinct spots on the 2-D gels in such experiments.

The chief disadvantages of 2-D gels are lack of reproducibility, failure to resolve most proteins greater than approximately 100 kDa, failure to routinely detect more than 1000 spots that can be identified by mass spectrometry, and the inability to separate most membrane proteins. The lack of dynamic range limits analysis to less than half of yeast proteins (68). This problem is even more serious in more complex organisms such as humans. Moreover, special gels are necessary to resolve extremely acidic or basic proteins. In spite of the advances in 2-D gel electrophoresis such as precast strips for immobilized pH gradient (IPG) and

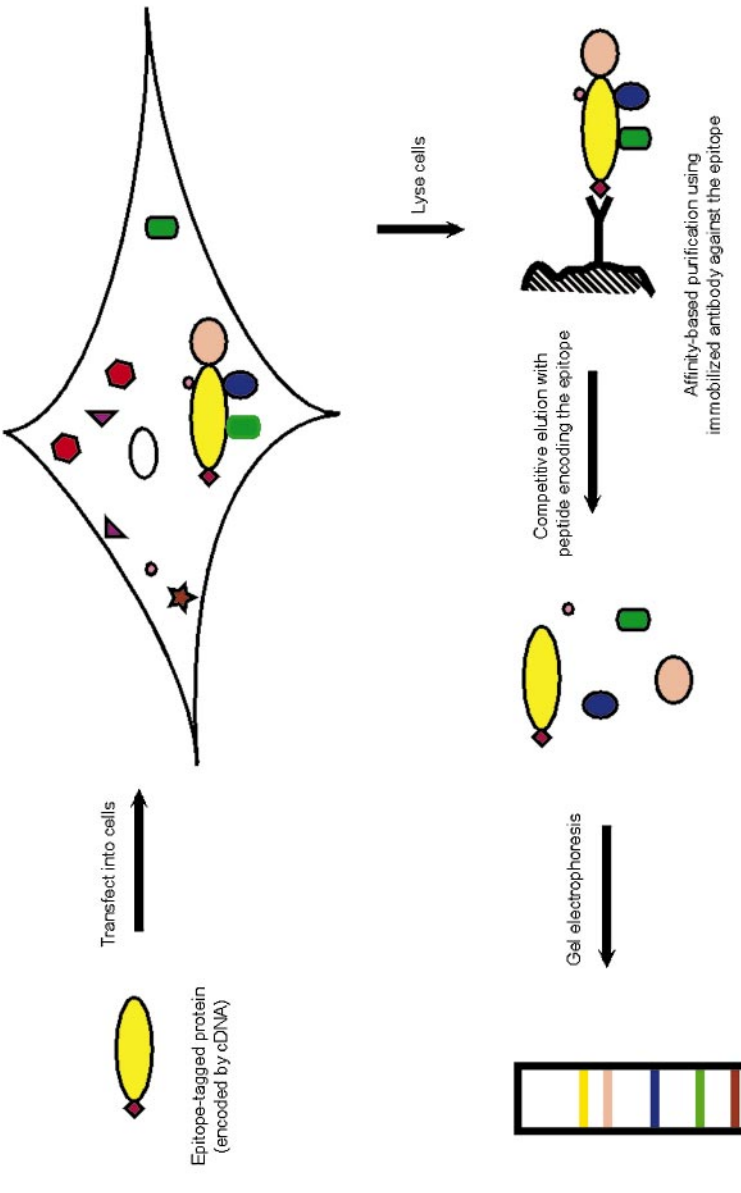
narrow pI range (zoom gels) strips for the first dimension, it is quite unlikely that the 2-D gels will be usable to compare and analyze highly complex samples in the foreseeable future. Currently, the DNA-based microarrays are quite useful and have been much more successful than 2-D gels at profiling the distinguishing components on any two samples such as tumor versus normal tissue (69–72).

Expression proteomics based completely on mass spectrometric methods [for example, the isotope-coded affinity tag (ICAT) strategy of Gygi et al] can in principle be completely automated and accurately identifies and quantitates a protein in a sample in the same step. Although this type of quantitative proteomics has to deal with the dynamic range problem as well, it will probably have a significant impact on areas such as toxicology where changes in the several hundred most abundant proteins are highly informative.

Protein Interactions and Protein Complexes

Functional proteomics attempts to correlate identification and analysis of proteins to the function of genes or proteins. Strategies outlined in this and subsequent sections are examples of functional proteomics. With the discovery of a variety of modular protein domains that have specific binding partners, it has become clear that most proteins occur in protein complexes and that understanding the function of a protein within the cell necessitates identification of its interacting partners (73). A number of purification strategies have been used to purify protein complexes. In the postgenomic era, it is important to use methods that use generic approaches in order to facilitate large-scale studies. Epitope-tagging of proteins and expressing them within cells followed by immunoprecipitation using epitope-specific antibodies is one such approach. Epitopes against which good antibodies exist include FLAG, HA, and Myc. It should be straightforward to establish optimal conditions for lysis and immunoprecipitation protocols depending on the protein under study. Once the protein is immunoprecipitated, the entire protein complex can be specifically eluted by a soluble peptide encoded for the epitope used (see Figure 10). A modification to this approach is to use two epitope tags (74). This may provide decreased binding to nonspecific proteins as well as improve the recovery of the protein complex. Besides these in vivo methods, binding assays can also be done in vitro. For this purpose, bacterially expressed GST

Figure 10 Tagging schematic. The figure shows various steps involved in the identification of interacting proteins by using an epitope-tagging strategy. The cDNA of interest is first cloned into a vector that provides an epitope tag. This is followed by transfection of the tagged "bait" into the cell of interest. The cells are then lysed and the lysates purified by affinity purification using an antibody against the epitope. Proteins bound specifically to the bait protein are eluted by competitive elution using a peptide that encodes the epitope. The proteins are then resolved by gel electrophoresis followed by mass spectrometric identification.



fusion proteins are generally used. Use of the GST moiety allows easy purification of proteins using glutathione bound to sepharose or agarose beads. It may also stabilize certain proteins that are easier to express as GST fusion proteins than as intact untagged proteins.

Alternative strategies have isolated components of other complexes such as the spliceosome complex (75), the Arp 2/3 complex (76), the p24 complex in the yeast (77), and many more. For example, in the case of the human spliceosome complex, RNA was used as bait (35), and in the case of the yeast spindle pole body and the nuclear pore complex, traditional biochemical fractionation was used (78, 79). It is now becoming possible to purify complete organelles to study their protein contents on a larger scale (80, 81).

Signal Transduction Pathways

Signal transduction pathways have been studied using a number of approaches including genetic, biochemical, and functional strategies. Biochemical approaches to identify signaling molecules have in the past been hampered by the fact that large amounts of pure protein are generally required for identification by Edman degradation. The ease of sequencing by mass spectrometry has revolutionized research in signal transduction pathways. Areas of biology that have been helped by proteomics range from apoptosis to cell division to cancer biology (see for example 82–85).

Receptor-mediated signaling pathways generally utilize Ser/Thr or tyrosine phosphorylation of cellular proteins to relay the signal from the membrane to the nucleus. Receptors possessing intrinsic tyrosine kinase activity such as epidermal growth factor (EGF) or platelet-derived growth factor (PDGF) receptors are therefore eminently suitable for a proteomic analysis. In this case, antiphosphotyrosine antibodies are used to enrich all of the tyrosine-phosphorylated substrates followed by one-dimensional electrophoresis to resolve proteins (see Figure 11). Bands that are found in growth-factor-treated but not untreated cells are simply excised and analyzed by mass spectrometry. Several components of the signal transduction pathway of the EGF receptor have been identified by such an approach (86, 87). The same strategy can also be applied to the global study of signaling pathways of T-cell, B-cell, and cytokine receptors. An alternative method is to resolve the cell lysates by 2-D gels and then Western blot to identify proteins that are phosphorylated and then to perform mass spectrometric identification. This was used to study the substrates of the PDGF receptor (88). The major drawback of this method is that unless the proteins are enriched, it is not easy to identify key signaling molecules because they are generally in low abundance. A study using mass spectrometry as a readout for studying proteins associated with the N-methyl-D-aspartate (NMDA) receptor was published recently (89).

Signaling pathways involving serine and threonine phosphorylation are not currently easy to study for the same reason. Antibodies or other agents to enrich for such serine- or threonine-phosphorylated substrates are not currently available. We

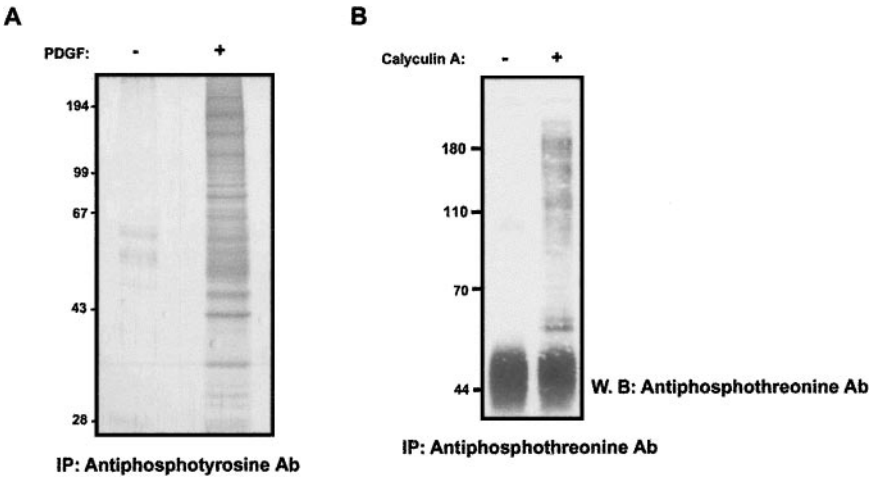


Figure 11 Analysis of phosphotyrosine and phosphothreonine phosphorylated proteins. (*A*) Purification of tyrosine-phosphorylated molecules in the signaling pathway of the PDGF receptor. The panel shows a silver-stained gel of proteins immunoprecipitated from NIH 3T3 fibroblasts left untreated (−) or treated (+) with PDGF using an antiphosphotyrosine antibody. (*B*) Purification of phosphothreonine-containing proteins. The panel shows antiphosphothreonine Western blotting of a gel to visualize phosphothreonine-containing proteins in immunoprecipitates using an antiphosphothreonine antibody from NIH 3T3 fibroblasts left untreated (−) or treated (+) with a serine/threonine phosphatase inhibitor, calyculin A.

have recently used some antiphosphothreonine antibodies to immunoprecipitate proteins that are phosphorylated when calyculin A, a Ser/Thr phosphatase inhibitor, is added. Our results suggest that in the future, a global strategy to study these pathways should be successful (Figure 11).

Secreted Proteins

Direct protein identification can be very useful in scenarios where other tools such as bioinformatics may be limited. Secreted proteins can be quite small (such as growth factors) or quite large (as seen in extracellular matrix components) in size. Most growth factors do not share extensive primary sequence homology and only the cysteine residues that specify folding of the proteins may be conserved. Furthermore, most growth factors are generally encoded by less than 150 amino acids. All of these factors make prediction of secreted proteins such as growth factors from the genome difficult by computational approaches. Work with 2-D gels to make a comprehensive map of secreted proteins from a number of cell lines has been initiated by Celis and coworkers (89a). We have successfully used 1-D gel electrophoresis to characterize secreted proteins during the process of adipogenesis (Figure 12). Such an approach may also be used to identify growth factors that are induced by cytokines and hormones.

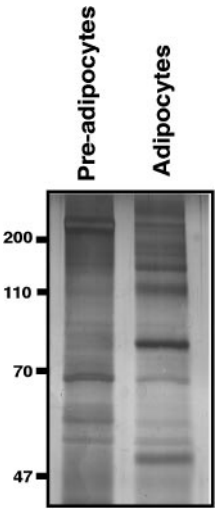
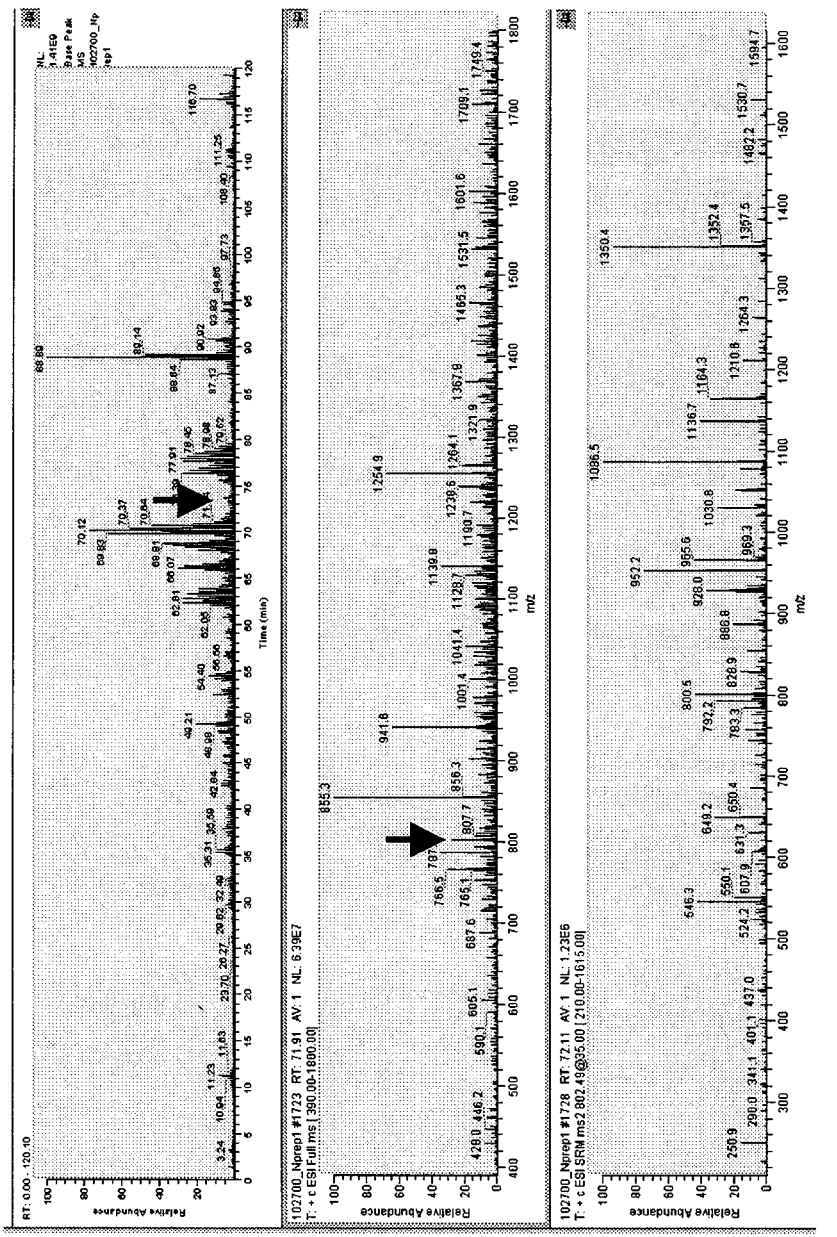


Figure 12 Direct proteomic analysis of secreted proteins. The figure shows a silver-stained gel of concentrated conditioned media from preadipocytes and adipocytes. The cells were grown normally in serum-containing medium, and in the case of adipocytes, allowed to differentiate using standard protocols. They were then washed extensively and allowed to grow in serum-free medium for approximately 18 hours before harvesting the conditioned medium. Several novel factors were identified.

Analyzing Protein Complexes and Organelles Without Gel Separation

As indicated above, multiprotein complexes can also be analyzed without gel separation of the proteins. Components of the yeast ribosome and the interchromatin granule cells have been studied in this manner (47, 90). Advantages of this approach include minimized sample handling and potentially very high sensitivity, as proteins can be digested in solution in small volumes, without loss of the released peptides in a gel matrix. Disadvantages are the lack of a quantitative overview of the whole set of proteins that the gel provides, as well as the lack of a molecular weight estimate of the protein. Furthermore, it is more difficult to address the degree of purification, which leads to the identification of abundant "household" proteins that may not be part of the complex under study (90, 91).

Figure 13 Analysis of a large protein structure by direct analysis of the unseparated proteins. Nucleoli from HeLa cells were purified and the protein component was subjected to tryptic digestion in solution. Part of the resulting crude peptide mixture was loaded onto a 75-μm column in an arrangement similar to that shown in Figure 7 and analyzed on an ion trap mass spectrometer. (*A*) The ion current of the most abundant peak at each MS spectrum. The mass spectrometer automatically extracts the masses of the five most abundant peaks, isolates them in turn, and obtains MS/MS spectra for each of them. More than 1000 peptides were fragmented in a single experiment. (*B*) An example of an MS spectrum at the position marked by the arrow in panel *A*. (*C*) The MS/MS spectrum of the ion marked by the arrow in panel *B*.



We have initiated a study of the components of the nucleolus in human cells and identified several hundred novel proteins by a combination of 1-D and 2-D electrophoresis and mass spectrometry as well as direct digestion and mixture analysis (see Figure 13) (CE Lyon, A Leung, JS Andersen, Y-W Lam, A Fox, M Mann & AI Lamond, unpublished data). Although mixtures from 1-D gels in these cases can be quite complex, knowledge of the protein molecular weight is preserved, helping in the analysis of novel gene products.

OUTLOOK

Mass spectrometry is the core technique of proteomics. Progress in instrumentation continues to be made at a fast pace. Automation will make it possible to obtain rates of data generation that will exceed those of genomics. This will allow the study of all protein complexes and organelles that can be purified. Many protein interactions can be studied by coimmunoprecipitation followed by mass spectrometric identification. Quantitative proteomics will most likely be achieved by stable isotope methods in combination with mass spectrometry.

Apart from the pressing areas of automation of data acquisition and interpretation, areas for future research will be the analysis of protein modification on a large scale. Cross-linking studies will tell us not only about the composition but also about the spatial organization of protein complexes. There is much room for creativity in connecting cell and molecular biological strategies with the powerful mass spectrometric capabilities to solve questions that could not previously be addressed.

ACKNOWLEDGMENTS

We are thankful for the help of and stimulating discussions with our colleagues at Protein Interaction Laboratory (PIL) and at Protana. Juri Rappsilber provided some of the figures. MM is grateful to Steven Gygi, Harvard Medical School, for a stimulating stay in his laboratory. Work at PIL is partially funded by a generous grant of the Danish National Research Foundation to the Center of Experimental BioInformatics (CEBI).

LITERATURE CITED

1. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. 1989. *Science* 246:64–71
2. Karas M, Hillenkamp F. 1988. *Anal. Chem.* 60:2299–301
3. Pandey A, Mann M. 2000. *Nature* 405:837–46
4. Klose J. 1975. *Humangenetik* 26:231–43
5. O'Farrell PH. 1975. *J. Biol. Chem.* 250: 4007–21
6. Vorm O, Roepstorff P, Mann M. 1994. *Anal. Chem.* 66:3281–87
7. Jensen ON, Podtelejnikov A, Mann M.

1996. *Rapid Commun. Mass Spectrom.* 10:1371–78
8. Krause E, Wenschuh H, Jungblut PR. 1999. *Anal. Chem.* 71:4160–65
9. Wilm M, Mann M. 1996. *Anal. Chem.* 68:1–8
10. Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, et al. 1996. *Nature* 379:466–69
11. Spengler B, Kirsch D, Kaufmann R, Jaeger E. 1992. *Rapid Commun. Mass Spectrom.* 6:105–8
12. Morris HR, Paxton T, Dell A, Langhorn J, Berg M, et al. 1996. *Rapid Commun. Mass Spectrom.* 10:889–96
13. Shevchenko A, Chernushevich I, Standing KG, Thompson B, Wilm M, Mann M. 1997. *Rapid Commun. Mass Spectrom.* 11:1015–24
14. Louris JN, Cooks RG, Syka JEP, Kelly PE, Stafford GC Jr, Todd JFJ. 1987. *Anal. Chem.* 59:1677–85
15. Kaiser RE Jr, Cooks RG, Syka JEP, Stafford GC Jr. 1990. *Rapid Commun. Mass Spectrom.* 4:30–33
16. Jonscher KR, Yates JR 3rd. 1997. *Anal. Biochem.* 244:1–15
17. Valaskovic GA, Kelleher NL, McLafferty FW. 1996. *Science* 273:1199–2202
18. Jensen PK, Pasa-Tolic L, Anderson GA, Horner JA, Lipton MS, et al. 1999. *Anal. Chem.* 71:2076–84
19. Shevchenko A, Loboda A, Ens W, Standing KG. 2000. *Anal. Chem.* 72:2132–41
20. Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, et al. 2000. *Anal. Chem.* 72:552–58
21. Shevchenko A, Wilm M, Vorm O, Mann M. 1996. *Anal. Chem.* 68:850–58
22. Pandey A, Andersen J, Mann M. 2000. *Science.* Online (STKE). http://www.stke.org/
23. Biemann K, Scoble HA. 1987. *Science* 237:992–98
24. Hunt DF, Shabanowitz J III, Yates JR 3rd, Zhu N-Z, Russell DH. 1987. *Proc. Natl. Acad. Sci. USA* 84:620–23
25. Berndt P, Hobohm U, Langen H. 1999. *Electrophoresis* 20:3521–26
26. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. 1999. *Electrophoresis* 20:3551–67
27. Eriksson J, Chait BT, Fenyo D. 2000. *Anal. Chem.* 72:999–1005
28. Clauser KR, Baker P, Burlingame AL. 1999. *Anal. Chem.* 71:2871–82
29. Jensen ON, Podtelejnikov AV, Mann M. 1997. *Anal. Chem.* 69:4741–50
30. Mann M, Wilm MS. 1994. *Anal. Chem.* 66:4390–99
31. Eng JK, McCormack AL, Yates JR 3rd. 1994. *J. Am. Soc. Mass Spectrom.* 5:976–89
32. Yates JR 3rd, McCormack AL, Schieltz D, Carmack E, Link A. 1997. *J. Protein Chem.* 16:495–97
33. Costanzo MC, Hogan JD, Cusick ME, Davis BP, Fancher AM, et al. 2000. *Nucleic Acids Res.* 28:73–76
34. Mann M. 1996. *Trends Biochem. Sci.* 21:494–95
35. Neubauer G, King A, Rappsilber J, Calvio C, Watson M, et al. 1998. *Nat. Genet.* 20:46–50
36. Vestal ML. 1990. In *Methods in Enzymol.* 193:107–30
37. Gale DC, Smith RD. 1993. *Rapid Commun. Mass Spectrom.* 7:1017–21
38. Wilm MS, Mann M. 1994. *Int. J. Mass Spectrom. Ion Processes* 136:167–80
39. Emmett MR, Caprioli RM. 1994. *J. Am. Soc. Mass Spectrom.* 5:605–13
40. Cox AL, Skipper J, Chen Y, Henderson RA, Darrow TL, et al. 1994. *Science* 264:716–19
41. Chervet JP, Ursem M, Salzmann JP. 1996. *Anal. Chem.* 68:1507–12
42. Davis MT, Stahl DC, Hefta SA, Lee TD. 1995. *Anal. Chem.* 67:4549–56
43. Gatlin CL, Kleemann GR, Hays LG, Link AJ, Yates JR 3rd. 1998. *Anal. Biochem.* 263:93–101
44. Shabanowitz J, Settlage RE, Marto JA, Christian RE, White FM, et al. 2000. In *Mass Spectrometry in Biology and*

Medicine, ed. AL Burlingame, SA Carr, MA Baldwin, pp. 163–77. Totowa, NJ: Humana Press
45. Martin SE, Shabanowitz J, Hunt DF, Marto JA. 2000. *Anal. Chem.* 72:4266–74
46. Davis MT, Lee TD. 1997. *J. Am. Soc. Mass Spectrom.* 8:1059–69
47. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, et al. 1999. *Nat. Biotechnol.* 17:676–82
48. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. 1999. *Nat. Biotechnol.* 17:994–99
49. Gatlin CL, Eng JK, Cross ST, Detter JC, Yates JR 3rd. 2000. *Anal. Chem.* 72:757–63
50. Whitelegge JP, le Coutre J, Lee JC, Engel CK, Prive GG, et al. 1999. *Proc. Natl. Acad. Sci. USA* 96:10695–98
51. le Coutre J, Whitelegge JP, Gross A, Turk E, Wright EM, et al. 2000. *Biochemistry* 39:4237–42
52. Hunter T. 1998. *Philos. Trans. R. Soc. London Ser. B* 353:583–605
53. The C. elegans Sequencing Consortium. 1998. *Science* 282:2012–18
54. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, et al. 2000. *Science* 287:2185–95
55. Affolter M, Watts JD, Krebs DL, Aebersold R. 1994. *Anal. Biochem.* 223:74–81
56. Zhang XL, Herring CJ, Romano PR, Szczepanowska J, Brzeska H, et al. 1998. *Anal. Chem.* 70:2050–59
57. Carr SA, Huddleston MJ, Annan RS. 1996. *Anal. Biotechnol.* 239:180–92
58. Betts JC, Blackstock WP, Ward MA, Anderton BH. 1997. *J. Biol. Chem.* 272:12922–27
59. Neubauer G, Mann M. 1999. *Anal. Chem.* 71:235–42
60. Qin J, Chait BT. 1997. *Anal. Chem.* 69: 4002–9
61. Posewitz MC, Tempst P. 1999. *Anal. Chem.* 71:2883–92
62. Nuwaysir LM, Stults JT. 1993. *J. Am. Soc. Mass Spectrom.* 4:662–69
63. Powell KA, Valova VA, Malladi CS, Jensen ON, Larsen MR, Robinson PJ. 2000. *J. Biol. Chem.* 275:11610–17
63a. Steen H, Küster B, Fernandez M, Pandey A, Mann M. 2001. *Anal. Chem.* 73:1440–48
64. Burlingame AL, Boyd RK, Gaskell SJ. 1998. *Anal. Chem.* 70:R647–716
65. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. 1999. *Proc. Natl. Acad. Sci. USA* 96:6591–96
66. Munchbach M, Quadroni M, Miotto G, James P. 2000. *Anal. Chem.* 72:4047–57
67. Page MJ, Amess B, Townsend RR, Parekh R, Herath A, et al. 1999. *Proc. Natl. Acad. Sci. USA* 96:12589–94
68. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. 2000. *Proc. Natl. Acad. Sci. USA* 97:9390–95
69. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, et al. 1999. *Science* 286:531–37
70. Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, et al. 2000. *Science* 287:873–80
71. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, et al. 2000. *Nature* 403:503–11
72. Richter J, Wagner U, Kononen J, Fijan A, Bruderer J, et al. 2000. *Am. J. Pathol.* 157:787–94
73. Pawson T, Scott JD. 1997. *Science* 278: 2075–80
74. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. 1999. *Nat. Biotechnol.* 17:1030–32
75. Neubauer G, Gottschalk A, Fabrizio P, Séraphin B, Lührmann R, Mann M. 1997. *Proc. Natl. Acad. Sci. USA* 94:385–90
76. Winter D, Podtelejnikov AV, Mann M, Li R. 1997. *Curr. Biol.* 7:519–29
77. Rowley A, Choudhary JS, Marzioch M, Ward MA, Weir M, et al. 2000. *Methods* 20:383–97
78. Wigge PA, Jensen ON, Holmes S, Soues S, Mann M, Kilmartin JV. 1998. *J. Cell Biol.* 141:967–77

79. Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, Chait BT. 2000. *J. Cell Biol.* 148:635–51
80. Bell AW, Ward MA, Freeman HN, Choudhary JS, Blackstock WP, et al. 2001. *J. Biol. Chem.* 276:5152–65
81. Peltier JB, Friso G, Kalume DE, Roepstorff P, Nilsson F, et al. 2000. *Plant Cell* 12:319–42
82. Muzio M, Chinnaiyan AM, Kischkel FC, Rourke KO, Shevchenko A, et al. 1996. *Cell* 85:817–27
83. Lingner J, Hughes TR, Shevchenko A, Mann M, Lundblad V, Cech TR. 1997. *Science* 276:561–67
84. Zachariae W, Shevchenko A, Andrews PD, Ciosk R, Galova M, et al. 1998. *Science* 279:1216–19
85. Ciosk R, Zachariae W, Michaelis C, Shevchenko A, Mann M, Nasmyth K. 1998. *Cell* 93:1067–76
86. Pandey A, Podtelejnikov AV, Blagoev B, Bustelo XR, Mann M, Lodish HF. 2000. *Proc. Natl. Acad. Sci. USA* 97:179–84
87. Pandey A, Fernandez MM, Steen H, Blagoev B, Nielsen MM, et al. 2000. *J. Biol. Chem.* 275:38633–39
88. Soskic V, Gorlach M, Poznanovic S, Boehmer FD, Godovac-Zimmermann J. 1999. *Biochemistry* 38:1757–64
89. Husi H, Ward MA, Choudhary JS, Blackstock WP, Grant SG. 2000. *Nat. Neurosci.* 3:661–69
89a. Celis JE, Ratz GP, Madsen P, Gesser B, Lauridsen JB, et al. 1989. *Electrophoresis.* 10:76–115
90. Mintz PJ, Patterson SD, Neuwald AF, Spahr CS, Spector DL. 1999. *EMBO J.* 18:4308–20
91. Verma R, Chen S, Feldman R, Schieltz D, Yates J, et al. 2000. *Mol. Biol. Cell* 11:3425–39

Annual Review of Biochemistry
Volume 70, 2001

CONTENTS