Mass Spectrometry in Proteomics

Ruedi Aebersold* and David R. Goodlett

The Institute for Systems Biology, 4225 Roosevelt Way NE, Seattle, Washington 98105

Received July 24, 2000

Contents

I. Introduction

Proteomics can be viewed as an experimental approach to explain the information contained in genomic sequences in terms of the structure, function, and control of biological processes and pathways. Proteomics attempts to study biological processes comprehensively by the systematic analysis of the proteins expressed in a cell or tissue. Mass spectrometry (MS) is currently proteomic's most important tool.

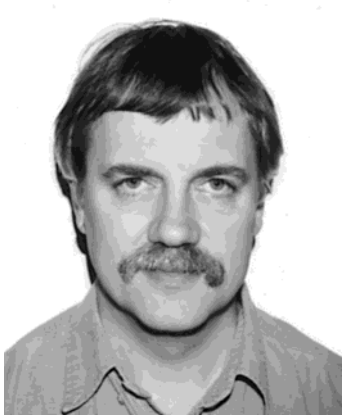
A. Genomics and Proteomics

The classical biochemical approach to study biological processes has been based on the purification to homogeneity by sequential fractionation/assay cycles of the specific activities that constitute the process; the detailed structural, functional, and regulatory analysis of each isolated component; and the reconstitution of the process from the isolated components. The Human Genome Project and other genome sequencing programs are turning out in rapid succession the complete genome sequences of specific species and thus, in principle, the amino acid sequence of every protein potentially encoded by that species.[1,2] It is to be expected that this information resource (unprecedented in the history of biology) will enhance traditional research methods such as the biochemical approach and also catalyze fundamentally different research paradigms, one of which is proteomics.[3−5]

The programs to sequence the entire human genome along with the genomes of a number of other species have been extraordinarily successful. The genomes of 46 microbial species (TIGR Microbial Database; www.tigr.org) have been completed, and the genomes of over 120 other microbial species are in the process of being sequenced. Additionally, the more complex genomes of eukaryotes, in particular those of the genetically well-characterized unicellular organism *Saccharomyces cerevisiae* and the multicellular species *Caenorhabditis elegans* and *Drosophila melanogaster* have been sequenced completely; a "draft sequence" of the rice genome has been published; and completion of the human and arabidopsis (92% complete in May 2000) genomes appear imminent.[6−11] Even in the absence of complete genomic sequences, rich DNA sequence databases have been publicly available, including those containing over 2.1 million human and over 1.2 million murine expressed sequence tags (ESTs).[12] ESTs are stretches of approximately 300−500 contiguous nucleotides representing partial gene sequences that are being generated by systematic single pass sequencing of the clones in cDNA libraries. On the time scale of most biological processes, with the notable exception of evolution, the genomic DNA sequence can be viewed as static. A genomic sequence database therefore represents an information resource akin to a library. Intensive efforts are underway to assign "function" to individual sequences in sequence databases. This is attempted by the analysis of linear sequence motifs or higher order structural motifs that indicate a statistically significant similarity of a sequence to a family of sequences with known function or by other means such as comparison of homologous protein functions across species.[13−17] These efforts will lead

* Corresponding author telephone: (206)732-1200; fax: (206)732-1299; e-mail: ruedi@systemsbiology.org.

10.1021/cr990076h CCC: $36.00
Published on Web 01/23/2001



Dr. Aebersold is a founding member of the Institute for Systems Biology in Seattle, WA, where he leads a research effort that is focused on developing new methods and technologies for understanding the structure, function, and control of complex biological systems. He completed his undergraduate studies in biology at the University of Basel, Switzerland, in 1979 and received a Ph.D. in Cell Biology at the University of Basel in 1984. Holding fellowships from the Swiss National Science Foundation and EMBO, he joined the California Institute of Technology as a postdoctoral fellow (1984–1986) and remained at Caltech as a senior research fellow (1986–1988). In 1988, he joined the University of British Columbia in Vancouver as an assistant professor in the Department of Biochemistry and Molecular Biology and as a senior investigator at the Biomedical Research Centre. In 1993, he moved to the University of Washington as an associate professor in Molecular Biotechnology and was promoted to full professor in 1998. In 2000, he left the University of Washington and joined the Institute for Systems Biology as co-founder and full faculty member. His research and teaching have been recognized by a long-term fellowship from the European Molecular Biology Organization (EMBO), by a scholarship from the Swiss National Science Foundation, by the Killam Research Prize, and by the Pehr Edman Award. He is a senior editor for the journal *Physiological Genomics,* has been a member of the Editorial Advisory Boards of *Protein Science* (1992–1998), *Functional Proteomics* (1999–present), *Analytical Biochemistry* (1991–present), *Functional and Integrative Genomics* (1999–present), and *Electrophoresis* (1989–1993).

Dr. Goodlett is a Senior Research Scientist and Director of the Proteomics Laboratory at the Institute for Systems Biology in Seattle, WA. Prior to this, he took an extended sabbatical (1998–1999) from the pharmaceutical industry to work with Prof. Ruedi Aebersold in the Department of Molecular Biotechnology at the University of Washington. During this time, he developed sensitive analytical methods for detection and sequencing of phosphopeptides. While employed in the pharmaceutical industry (1993–1997), he carried out analytical research on an HIV therapeutic for Johnson & Johnson, Inc. and drug discovery research in the field of immunology for Bristol-Myers Squibb, Inc. He is an expert in the field of protein/peptide characterization by mass spectrometry and in microscale separation sciences. Dr. Goodlett was a NORCUS Postdoctoral Fellow in the laboratory of Richard D. Smith at Battelle-Memorial Institute (1991–1992) where he developed mass spectrometric methods for determination of the thermodynamic properties of protein–protein complexes and methods to increase the sensitivity of capillary electrophoresis-mass spectrometry. He obtained a Ph.D. in Biochemistry from North Carolina State University (1991) in the protein mass spectrometry laboratory of Richard B. van Breemen and holds M.S. (1988) and B.S. (1982) degrees in Chemistry from Auburn University.

to a more richly annotated sequence database and, not by themselves, to an explanation of the structure, function, and control of biological processes.

The proteome has been defined as the protein complement expressed by a genome.[18−21] This somewhat restrictive definition implies a static nature of the proteome. In reality, the proteome is highly dynamic; the types of expressed proteins, their abundance, state of modification, subcellular location, etc. being dependent on the physiological state of the cell or tissue. Therefore, the proteome reflects the cellular state or the external conditions encountered by a cell, and proteome analysis can be viewed as a genome-wide assay to differentiate and study cellular states and to determine the molecular mechanisms that control them.[22] Considering that the proteome of a differentiated cell is estimated to consist of thousands to a few ten-thousands of different types of proteins with an estimated dynamic range of expression of at least 5 orders of magnitude, the prospects for proteome analysis appear daunting. However, the availability of (genomic) DNA databases listing the sequence of every potentially expressed protein and rapid advances in technologies capable of identifying the proteins that are actually expressed now make proteomics a realistic proposition. MS is one of the essential legs on which current proteomics technology stands.

B. MS and Proteomics

During the decade of the 1990s, changes in MS instrumentation and techniques revolutionized protein chemistry and fundamentally changed the analysis of proteins. These changes were catalyzed by two technical breakthroughs in the late 1980s: the development of the two ionization methods electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI).[23−25] These methods solved the difficult problem of generating ions from large, nonvolatile analytes such as proteins and peptides without significant analyte fragmentation. Because of the lack or minimal extent of analyte fragmentation during the ESI and MALDI processes, they are also referred to as “soft” ionization methods. In fact they are so soft that under specific conditions even noncovalent interactions may be maintained during the ionization process. ESI gained immediate popularity because of the ease with which it could be interfaced with popular chromatographic and electrophoretic liquid-phase separation techniques and quickly supplanted fast atom bombardment as the ionization method of choice for protein and peptide samples dissolved in a liquid phase.[26] Furthermore, due to the propensity of ESI to produce multiply charged analytes, simple quadrupole instruments and other types of mass analyzers with limited m/z range could be used to detect analytes with masses exceeding the nominal m/z range of the instrument. For different but no less compelling reasons, MALDI also rapidly gained popularity. The

time-of-flight (TOF) mass analyzer most commonly used with MALDI is robust, simple, and sensitive and has a large mass range. MALDI mass spectra are simple to interpret due to the propensity of the method to generate predominantly singly charged ions. The method is relatively resistant to interference with matrixes commonly used in protein chemistry.

Direct measurement of the molecular weight of large [>10 000 mass units (u)] polypeptides was quickly demonstrated with both ESI and MALDI.[25,27] More recently, the mass determination of very large proteins in excess of 100 000 u has been realized.[28,29] While the early roles of ESI-MS and MALDI-MS in protein analysis were essentially those of accurate balances,[30] the ease with which proteins and peptides could be ionized by these methods rapidly made MS a complementary technique to nuclear magnetic resonance, X-ray crystallography, circular dichroism, and the classical methods of protein chemistry to study diverse aspects of protein structure and function. Numerous reports document the success MS has enjoyed in studies into the four structural classifications of proteins, namely, the *primary* structure or linear sequence of amino acids, the *secondary* structure or the folding of stretches of amino acids into defined structural motifs, the *tertiary* structure or the overall three-dimensional fold, and the *quaternary* structure or the spatial arrangement of folded polypeptides in multiprotein complexes. The application of MS to proteomics, the subject of this review, has to date been realized mostly for the study of protein primary structures. However, the following anecdotal examples hint at an increasing role of MS in the systematic study of protein higher order structures, i.e., structural proteomics, as well as of protein−ligand interactions.

Because of their relative softness of ionization, ESI and MALDI have been used in attempts to generate gas-phase ions of noncovalently associated, apparently intact protein complexes for the purpose of studying these structures by MS.[31] While controversy continues over whether this is a general approach applicable to all noncovalent complexes, there are documented cases in which information gained by gas-phase examination of protein−protein interactions appears to correlate well with data obtained from the same complexes in the liquid phase. An example is the enzyme ribonuclease S that requires a noncovalent association of a peptide and a protein for catalytic activity. Measurements of thermal denaturation of this complex in the gas phase and in solution indicated that the enthalpy of dissociation as determined by ESI-MS (gas phase) correlated well with measurements made in solution by calorimetry.[32] Other early studies showed a direct correlation between the individual observed relative abundances for a series of enzyme−inhibitor (E−I) complexes in the gas phase and their ranking or affinity found by calculating K_d values from traditional kinetic data.[33] The use of MS for the analysis of noncovalent protein complexes has been competently reviewed and is not further discussed in this paper.[31,34]

Other studies focused on the use of hydrogen−deuterium (H−D) exchange to examine higher order structural features of proteins by MS.[35] These experiments are based on the assumption that not all of the exchangeable hydrogens in a protein exchange at the same rate and that the rate of exchange is an indicator of structural properties of a protein. Examples of structural features that can be analyzed in this way include solvent accessibility, based on the observation that solvent-exposed hydrogens exchange more rapidly than those shielded from solvent access, and hydrogen bonding, based on the observation that hydrogens involved in hydrogen bonds exchange at a slower rate than those not involved in hydrogen bonds. With these concepts in mind, Anderegg and co-workers used ESI-MS to study the transition of a peptide from α-helical to a denatured configuration.[36,37] Even simpler experiments without H−D exchange have shown that ESI-MS can be used to monitor the transition of a protein in solution from native to denatured state. Such experiments rely on the empirical observation that ESI mass spectra of proteins known to be unfolded in solution indicate a higher charge state (a greater number of protons associated with the protein) than the identical protein not subject to denaturation.[38−41] Thus, the transition of a protein from a folded to a denatured state can be followed by ESI-MS by examining the charge state distribution of the protein molecular ions. Other, perhaps less controversial, applications of MS to study protein higher order structure include the identification of spatially juxtaposed amino acids by chemical cross-linking or the determination of the extent of heavy atom incorporation prior to X-ray diffraction of protein crystals.[42,43] The use of cross-linkers of a defined length that are chemically reactive to specific amino acid side chains such as the primary amine of lysine have been used to examine both intra- and inter-protein distances. A study of the yeast nuclear pore complex is a recent example of this approach.[44]

MS, in particular the application of ESI coupled on-line with high-performance separation techniques such as capillary electrophoresis (CE) and HPLC, has had a dramatic effect on the sensitivity and the speed with which the primary structure of proteins and peptides can be determined. Advances in separation techniques, particularly their implementation in miniaturized formats on-line with high-performance mass spectrometers,[45−50] and the development of miniaturized sprayers as ESI ion sources[51−53] have reduced the amount of peptide required for complete and routine sequence characterization from several picomoles of peptide[54,55] in the mid-1980s to a few femtomoles and below by the mid-1990s.[56−59] The development of mass spectrometric techniques of yet higher throughput and sensitivity is an essential component of the emerging field of proteomics and is still forcefully pursued today. Through incremental improvements in on-line separation methods, sensitivities in the sub-femtomole peptide detection and identification range have been achieved with commercially available ion trap mass spectrometers.[60,61] For simple mass measurement, sensitivities into the

Table 1. Sources for MS-Based Protein Identification Tools

sponsor (application)	uniform resource locator (URL)
Eidgenossische Technische Hochschule (MassSearch)	http://cbrg.inf.ethz.ch
European Molecular Biology Laboratory (PeptideSearch)	http://www.mann.emblheidelberg.de
Swiss Institute of Bioinformatics (ExPASy)	http://www.expasy.ch/tools
Matrix Science (Mascot)	http://www.matrixscience.com
Rockefeller University (PepFrag, ProFound)	http://prowl.rockefeller.edu
Human Genome Research Center (MOWSE)	http://www.seqnet.dl.ac.uk
University of California (MS-Tag, MS-Fit, MS-Seq)	http://prospector.ucsf.edu
Institute for Systems Biology (COMET)	http://www.systemsbiology.org
University of Washington (SEQUEST)	http://thompson.mbt.washington.edu/sequest

zeptomole (10^{-21} mol) range have been observed with prototype FT-ICR-MS instruments.[62] The use of microfabricated devices connected on-line with ESI-MS[56,60,61,63] offers the exciting possibility of generating integrated analytical systems for sample manipulation, preparation, and analysis that promise to operate at unprecedented levels of automation, sensitivity, and throughput.

In this paper, we will review current methods that are the basis for proteome analysis, specifically mass spectrometric strategies for protein identification from biological matrixes, computational approaches for searching sequence databases, determination of protein expression levels, and characterization of phosphoproteins and phosphopeptides. We attempt to cover the most popular and promising techniques, but the review does not claim to be comprehensive. Furthermore, the mass spectrometers used for such studies, mainly the established MALDI time-of-flight[25] (TOF) mass spectrometer, the ESI-triple quadrupole[64,65] (TQ) mass spectrometer, the ESI ion trap[66,67] (IT) mass spectrometer, and the increasingly popular ESI-quadrupole-TOF [68−70] (Q-TOF), the ESI-TOF,[71] the MALDI-IT-TOF,[72] MALDI-TOF-TOF,[73] Fourier transform ion cyclotron resonance[74,75] (FT-ICR), MALDI-ion trap,[76,77] ESI-ion mobility,[78] and their respective operations, will not be described in detail here because this is the subject of another paper in this issue and other recent reviews.[79,80]

II. Methods for Protein Identification

Traditionally, proteins have been identified by de novo sequencing, most frequently by the automated, stepwise chemical degradation (Edman degradation) of proteins or isolated peptide fragments thereof.[81,82] These partial sequences were occasionally used to assemble the complete protein sequence from overlapping fragments but more frequently for the generation of probes for the isolation of the gene coding for the protein from a gene library. With the growing size of sequence databases, it became apparent that even relatively short and otherwise imperfect sequences (gaps, ambiguous residues) were useful for the identification of proteins. This was done by correlating information obtained experimentally from the analysis of peptides with sequence databases. The concept of identifying proteins by correlating information extracted from a protein or peptide with sequence databases rather than by de novo sequencing was significantly enhanced when it was realized that mass spectrometers were ideally suited to generate the required data. Furthermore, the methods initially developed for the isolation of small amounts of proteins and peptides for Edman sequencing were directly compatible with peptide analysis by LC-MS and LC-MS/MS. This further accelerated the implementation of mass spectrometric methods for protein identification.[21,82−88] Correlation of mass spectrometric data with sequence databases also depended on the development of novel search algorithms, a number of which are available on the worldwide web and listed in Table 1. Such algorithms use readily available constraints in a decision-making process that distinguishes the correct match from all other sequences in the database. The availability of complete sequence databases, the development of mass spectrometric methods, and the sequence database search algorithms therefore converged into a mature, robust, sensitive, and rapid technology that has dramatically advanced the ability to identify proteins and constitutes the basis of the emerging field of proteomics. In the following, we discuss the different approaches that have been developed for the identification of proteins by sequence database searching using data predominantly generated by MS.

A. Protein Identification Using Multiple Related Peptides

The methods described in this section use information obtained from analysis of multiple fragments of a single protein for database searching. Since the source of a specific fragment can only be unambiguously determined if a single, homogeneous protein is being analyzed, these methods require that proteins are highly purified. In the context of proteomics, a high degree of purification of multiple (ideally all) proteins in a sample is typically achieved by high-resolution two-dimensional gel electrophoresis (2DE). Therefore, these identification methods in conjunction with 2DE form the basis for many proteome projects.[89−92]

1. Nonmass Spectrometric Methods

As the subject of this paper is MS and proteomics, these methods are only mentioned peripherally. A more extensive treatment of nonmass spectrometric methods for use in proteome analysis can be found in a review by Wilkins et al.[18] It has been well-known for a long time that proteins differ considerably not only in their amino acid sequence but also in their amino acid composition.[93,94] Wilkins and co-workers therefore attempted to implement high throughput identification of proteins separated by 2DE by determining the accurate amino acid composition of

specific spots and submitting the data to a database search algorithm they developed.[95] It turned out that unambiguous protein identification by the amino acid composition alone was not always achieved and that secondary search constraints such as the isoelectric point and the molecular mass of the parent protein (as obtained from the spot coordinates in the 2D gel) were useful to increase the confidence of the search results and in many cases essential to obtain unambiguous identification. The method is quite sensitive to contaminating proteins present in the sample, either comigrating proteins or other contaminants, and has been essentially supplanted by mass spectrometric methods. Still the additional information provided by amino acid composition or even partial amino acid composition can be of value as an additional constraining parameter in sequence database searches. This was shown recently in a report that combined mass mapping (discussed below) with vapor-phase acid hydrolysis that was specific for protein cleavage after serine, threonine, aspartic acid, and glycine.[96]

2. Mass Spectrometric Methods

a. Principle of Peptide Mass Mapping. Peptide mass mapping is based on the insight that the accurate mass of a group of peptides derived from a protein by sequence-specific proteolysis (i.e., a mass map or fingerprint) is a highly effective means of protein identification. The principle behind protein identification by mass mapping is therefore quite simple conceptually and was implemented by several groups independently at approximately the same time.[85,97−100] Proteins of different amino acid sequence (Figure 1A) will, after proteolysis with a specific protease, produce groups of peptides the masses of which constitute mass fingerprints unique for a specific protein (Figure 1B). Therefore, if a sequence database containing the specific protein sequence is searched using selected masses (i.e., the observed peptide mass fingerprint), then the protein is expected to be correctly identified within the database. In the example shown in Figure 1B, the four peptide monoisotopic masses shown are sufficient to identify the protein as myoglobin and the species as *Equus caballus*. Various methods that automate this process have been developed and reviewed.[85] They vary in specific details but share the following sequence of steps: (i) Peptides are generated by digestion of the sample protein using sequence-specific cleavage reagents that allow residues at the carboxyl- or amino-terminus to be considered fixed for the search. For example, the enzyme trypsin that is popular for mass mapping leaves arginine (R) or lysine (K) at the carboxyl-terminus (Figure 1), and the N-termini of tryptic peptides (except for the N-terminal one) are expected to be the amino acid following a K or R residue in the protein sequence. (ii) Peptide masses are measured as accurately as possible in a mass spectrometer. An increase in mass accuracy will decrease the number of isobaric peptides for any given mass in a sequence database and therefore increase the stringency of the search. (iii) The proteins in the database are "digested" in silico using the rules that apply to the proteolytic method used

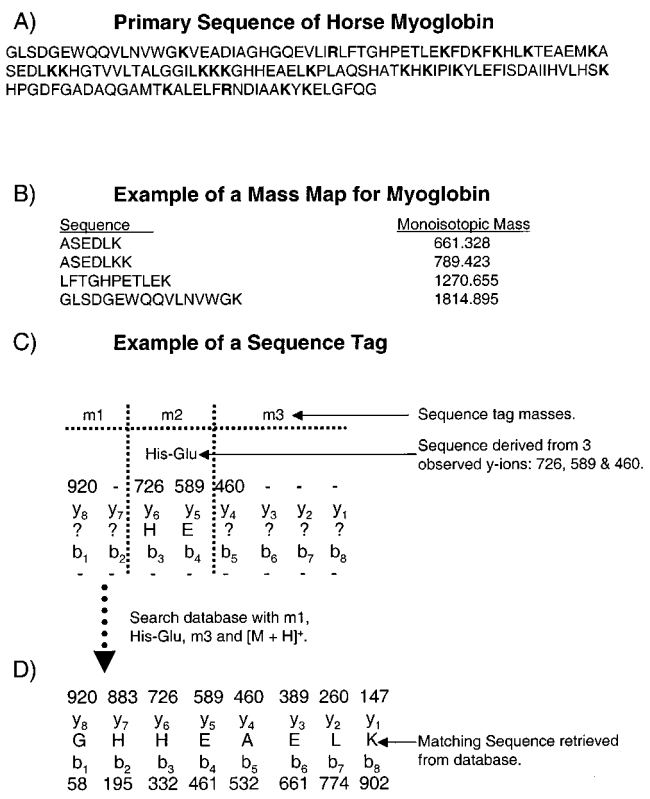


Figure 1. Protein identification. (A) Primary amino acid sequence of horse heart myoglobin with trypsin recognition sites (R and K) in bold. (B) Randomly selected monoisotopic mass values that serve as an example of a mass map. No simple rules exist for predicting all of the peptides that will be generated by digestion of a protein with trypsin. (C) One of the peptides from panel A presented as an example of the sequence tag method for protein identification. (D) Peptide fragment ion nomenclature and sequence of peptide from database determined with sequence tag information.

in the experiment to generate a list of theoretical masses that are compared to the set of measured masses. (iv) An algorithm is used to compare the set of measured peptide masses against those sets of masses predicted for each protein in the database and to assign a score to each match that ranks the quality of the matches. Obviously, for a protein to be identified its sequence has to exist in the sequence database being used for comparison. Both protein and DNA sequence databases are equally suited. If DNA sequence databases are being used, the DNA sequences are translated into protein sequences prior to digestion. The approach is therefore best suited for genetically well-characterized organisms where either the entire genome is known or extensive protein or cDNA sequence exists.

b. Pitfalls and Limitations of Mass Mapping. Protein identification by peptide mass mapping depends on the correlation of several peptide masses derived from the same protein with corresponding data calculated from the database. For this reason the method is suited neither for searches of EST databases nor for identification of proteins in complex mixtures if unseparated mixtures are proteolyzed. ESTs present a problem because they only represent a portion of a gene's coding sequence. Such segments may not be long enough to cover a sufficient number

of peptides observed in the mapping experiment to allow an unambiguous identification. Digests of unseparated protein mixtures present a problem for mass mapping because it is not apparent which peptides in the complex peptide mixture originate from the same protein. The mass mapping method is therefore most popular for the identification of proteins from microbial species for which complete genome sequences have been determined and for use with protein purification by 2DE where ancillary information on protein molecular weight and isoelectric point information can be used to aid identification. It is often combined with tandem MS of peptides (discussed later in this paper) in an iterative approach where as much information as possible is extracted by mass mapping, and this is followed by tandem MS to resolve the identification of any ambiguous remaining masses.

If a pure protein is digested and the resulting peptide masses are compared with the list of peptide masses predicted for that protein, two observations are typically made. First, not all of the predicted peptides are detected. Second, some of the measured peptide masses are not present in the list of masses predicted from the protein. The first problem, the missing masses, is usually due to a number of problems that can occur both before and during mass spectrometric analysis such as poor solubility, selective adsorption, ion suppression, selective ionization, very short peptide length, or other artifacts that cause sample loss or make specific peptides undetectable by MS. In rarer cases, e.g., in situations of alternative gene splicing, missing peptide masses may contain meaningful biological information. Unfortunately, it is not possible to distinguish between trivial and meaningful missing masses without further experimentation. Since a relatively low number of peptide masses are sufficient for the positive identification of a protein, missing peptide masses are not generally considered a problem. In contrast, unassigned peptide masses are a significant problem for protein identification by mass mapping and probably the single biggest source of misidentifications or missed identifications. Thus, to ensure that mass mapping results are reliable, it is important to understand the possible reasons for unassigned masses and to learn how to deal with them.[85,101−103] Unassigned masses may be observed for one or more of the following reasons: (i) Changes in the expected peptide masses by post-translational modification (e.g., phosphorylation adds a net 80 u to an amino acid mass), artifactual modifications arising from sample handling (such as oxidation of methionine), or post-translational processing (e.g., amino- or carboxyl-terminal processing). Some of these changes can be anticipated and incorporated into the search algorithm. (ii) Low fidelity proteolysis due to the presence of contaminating proteases that produce peptides unanticipated by the search algorithm (e.g., the presence of chymotryptic activity in a trypsin preparation) or missed cleavage sites. Again, this can be anticipated to some degree by the search algorithms. (iii) The presence of more than one protein in the sample. It needs to be stressed that bands in SDS gels frequently and spots in 2D gels occasionally contain more than one protein, even if the respective features appear concise and sharp. In some cases, additionally present proteins can be detected by iterative database searching with the masses left unassigned to the primary target protein. Keratins and other common proteins represent another source of protein contamination. (iv) The identified protein actually matches a sequence homologue or splice variant of that reported in the database. This must be confirmed using the sequence of genetically well-characterized species.[104] (v) The protein is misidentified (i.e., false-positive).

In this context, the specificity of the enzymes employed for protein digestion should be discussed in more detail. Obviously, the higher the fidelity of the enzyme in hydrolyzing peptide bonds, the more reliably the search can be done with a fixed amino- or carboxyl-terminus. Unfortunately, proteases are far from perfect enzymes. In addition to cleavage at expected amino acid residues, they tend to cleave at unexpected sites and tend to skip anticipated cleavage sites. The frequent observation that the protease products are not limited to the ones predicted from the expected enzymatic recognition sites is often due to contaminating protease activity but may also be due to a post-translational modification juxtaposed to the recognition site that blocks access by the enzyme. Furthermore, even highly purified trypsin appears to cleave at sites other than carboxyl-terminal to the expected recognition sites, K and R. The so-called "missed" cleavages produce "ragged" termini when two or more consecutive amino acids in a protein sequence are recognition sites for the enzyme. For example, when trypsin hydrolyzes myoglobin around the following sequence, ASEDLKKHGTVVTALGGILK (Figure 1A), it is expected that four peptides could be produced: ASEDLK, ASEDLKK, HGTVVTALGGILK, and KHGTVVTALGGILK. If this problem is anticipated, algorithms can be programmed to accommodate missed cleavages by allowing a given number to be entered as a parameter. Furthermore, the success of proteases to cleave proteins is dependent on accessibility to open stretches of primary amino acid sequence, and the native three-dimensional structure of the substrate protein will block access to many sites. Thus, proteins in solution are frequently not completely proteolyzed until they are denatured. Proteins separated by denaturing gel electrophoresis methods such sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) are highly accessible to proteases and generally yield rich and reproducible peptide maps. In addition to the SDS that coats the protein and forces native interactions apart, reducing compounds such as dithiothreitol are used to chemically break disulfide bonds further opening the protein structure.

c. Secondary Parameters for Enhanced Peptide Mass Mapping. It should not be surprising that the single most critical experimental parameter for protein identification by mass mapping is the accuracy of the peptide mass measurement.[105−108] For any single measured peptide mass, the list of isobaric masses found in the database will decrease as the

mass accuracy of the measurement increases. Increased mass accuracy not only increases search speeds but also increases reliability in the score.[102,106,108,109] However, even in cases in which highly accurate mass measurements are possible, it is frequently desirable to use ancillary information to confirm true positives and to eliminate false-positive matches. Such information can be factored into the search algorithms to further constrain a search. The value of such ancillary information for peptide mass mapping has been extensively discussed,[105] and the power of this approach was illustrated in a recent report in which the presence of the relatively rare amino acid cysteine was used, along with other readily available constraints, to identify a protein within the yeast genomic sequence database from the accurate mass of a single cysteine-containing peptide.[110]

Good constraints should be easily obtainable experimentally and be highly discriminating. Quite a large number of viable experimental approaches for the generation of search constraints to enhance peptide mass mapping have been developed. They have the common objectives of deducing the presence of specific amino acids in the peptide analyzed and/or the location of a specific amino acid within the peptide without the use of tandem mass spectrometry. They differ in the way these objectives are pursued. The presence of specific amino acids in a peptide has been detected by site-specific chemical modifications. The number of modifications induced and therefore the number of the specifically targeted residues in a peptide is determined from the mass differential of a peptide before and after the modification reaction. Methyl esterification, which adds +14 u for each carboxyl group (i.e., side chains of aspartic acid, glutamic acid, or the carboxyl-terminus),[111] and iodination of tyrosine, which adds +126.9 u for each tyrosine,[112] are examples of such reactions. Similarly, the presence of cysteinyl residues was detected by isotopic labeling of sulfhydryl groups using a 1:1 mixture of acrylamide and deuterated acrylamide,[113] and the partial amino acid composition of peptides was determined by measuring the number of exchangeable hydrogens via hydrogen−deuterium exchange.[114] To locate specific residues within the peptide sequence, a single step of Edman degradation of unseparated peptide mixtures[115] or aminopeptidase[116] treatment have been used to identify the amino-terminal residue. Secondary proteolytic digestions (or a parallel digestions) using an enzyme with a different specificity from the one used for the primary digestion have been used to locate specific residues within a peptide.[111,114] The carboxyl-terminal residue(s) were identified by carboxypeptidase treatment of peptides or peptide mixtures.[116,117] In cases in which gel separated proteins are being identified, properties deduced from the position of the protein in the gel (protein mass for SDS−PAGE and protein mass and pI for 2DE) are also frequently used to confirm the identity protein analyzed.

d. Generation of Data for Peptide Mass Mapping. Data for use with peptide mass mapping are commonly obtained via MALDI-TOF analysis. However, any mass spectrometer capable of generating mass accuracies around 100 ppm or better at 1000 u (i.e., 100 parts in 1000.000 or in the example accurate to the first place past the decimal point), in particular ESI-TOF and FT-ICR instruments, can be used to generate a mass map. For MALDI, analytes are spotted onto a metal plate either one at a time or, in a higher throughput format, multiple samples on the same plate. The samples are usually tryptic digests from proteins separated by 2DE, although proteins purified by other separation methods are also compatible with the method. Before deposition of the analytes, the matrix is placed on the plate or mixed in with the sample. The matrix will absorb energy from the laser causing the analytes to be ionized by MALDI. The m/z ratio of the ions is then typically measured based on the flight time in a field-free drift tube (as opposed to ion mobility MS where a field pushes ions through a gas) that constitutes the heart of the time-of-flight mass (TOF) analyzer. Using internal calibration on monoisotopic masses, a mass accuracy of 5 ppm at 1000 u can be achieved. An additional bonus for samples isolated from biological sources is that MALDI is compatible with biological buffers such as phosphate and Tris and low concentrations of urea, nonionic detergents, and some alkali metal salts. Peptide m/z ratios are calculated based on the energy equation ion $E = 1/2mv^2$ that accounts for contributions from kinetic energy, mass, and velocity. At a constant energy, low molecular weight ions will travel faster than high molecular weight ions—flight times of ions are inversely proportional to the square root of their molecular mass.

An inherent problem with the MALDI process is the small spread of kinetic energy that occurs during ionization. The spread reduces the resolving power and prevents the observation of the natural isotope distribution, even of small peptides. Two approaches, an ion mirror (reflectron) and "time-lag focusing" (a.k.a. delayed extraction), have been implemented in commercial instruments to overcome this problem. A reflectron is a device located at the end of the flight tube opposite from the ion source that decelerates the ions and then re-accelerates them back out of the reflectron toward a second detector. This is achieved by applying a decelerating voltage that is slightly higher than the accelerating voltage at the source. It has been observed that ions of lower kinetic energy do not penetrate as far into the reflectron as those of higher energy. Consequently, deeper penetrating high-energy ions can catch up, thereby decreasing the initial energy spread. The second approach to correct the initial spread of kinetic energies during MALDI is the time-lag focusing technique initially developed by Wiley and McLaren in 1953 and more recently reintroduced as "delayed extraction".[118,119] In this method, the MALDI ions are created in a field-free region and allowed to spread out before the extraction voltage is applied to accelerate them for their flight through the drift tube. This results in a significantly decreased energy spread of ions and thus higher resolution. Delayed extraction also limits peak broadening due to metastable decomposition from ions colliding in the source during continuous ion extrac-

tion. The effects of these improvements are significant. Delayed extraction can increase the mass resolution to ~2000−4000 for peptides in a linear instrument and, if combined with a reflectron instrument resolution, can further increase to ~3000−6000.[119,120]

e. Examples of Proteome Projects by Peptide Mass Mapping. *Haemophilus influenzae* has been the subject of one of the most extensive proteome efforts based on 2DE and mass mapping to date.[92] Soluble proteins were separated by 2DE. To enhance the separation range, immobilized pH gradient strips of various pH ranges[121] and second dimension SDS gels with different acrylamide concentrations and electrophoresis buffers with different trailing ions were used. Low-copy-number proteins were visualized by employing a series of protein extraction and chromatographic steps that included heparin chromatography, chromatofocusing, and hydrophobic interaction chromatography. Cell envelope-bound proteins were separated electrophoretically, either by immobilized pH gradient strips or a two-detergent system in which a cationic detergent was used in the first dimension and an anionic detergent in the second. A combination of MALDI-TOF MS and amino acid composition analysis identified a total of 502 proteins out of a genome of approximately 1742 ORFs.

In contrast or in addition to cataloguing the proteins expressed in a particular cell or tissue, global protein expression profiles, if analyzed quantitatively, can be useful to differentiate cell types or the same cell type in different physiological or pathological states. A recent study used 2DE to better examine the taxonomic relationship between several divergent yeast species. Yeast strains used in the brewing industry are known to be hybrid strains of at least two different genomes.[92] By analyzing the proteins expressed by three commonly used brewing strains by 2DE and comparing the resulting patterns to the patterns and identified protein spots for *S. cerevisiae*, it was established that *S. carlsbergensis*, *S. monacensis*. and *S. pastorianus* represented two divergent evolutionary patterns. One pattern originated from an *S. cerevisiae*-like genome and the other from *S. pastorianus* strain NRRL Y-1551. Numerous additional proteome projects have been attempted or are in progress (see current information at www.expasy.ch). Generally, peptide mass mapping is chosen as the method of choice for protein identification if a complete genomic sequence database is available. Protein identification using tandem mass spectrometry (see below) has proven to be superior however for the identification of proteins from species with large and incompletely sequenced genomes.

B. Protein Identification Using Single Peptides

Different amino acid compositions and permutations of an amino acid sequence can result in isobaric peptides. The amino acid sequence of a peptide is therefore more constraining than its mass for protein identification by sequence database searching.[122] At the mass accuracy achieved with the MALDI-TOF mass spectrometers that are frequently used for peptide mass measurement (10−100 ppm), several peptide masses from the same protein are required for unambiguous identification, whereas the amino acid sequence of even a relatively small peptide can uniquely identify a protein. The number of proteins that are targeted for identification in typical proteomics projects by far exceed the capacity of the traditional chemical sequencing methods, and mass spectrometric methods for the rapid generation of amino acid sequence information from intact proteins, while exciting, still require further development.[122−126] Large-scale protein identification therefore critically depends on tandem mass spectrometry for the generation of sequence-specific spectra for peptides. In this section, we describe the methods used for the generation of sequence-specific peptide mass spectra and their use for large-scale protein identification. These methods were initially designed to identify purified proteins (typically homogeneous protein spots in 2D gels) faster, more sensitively, and more reliably. More recently, essentially the same methods have been applied to also identify proteins in complex mixtures.[127]

1. Protein Identification via Sequence-Specific Peptide Mass Spectra

Tandem mass spectrometers and, to a more limited extent, single-stage mass spectrometers have the ability to fragment peptide ions and to record the resulting fragment ion spectra. For tandem mass spectrometers such as triple quadrupole, ion trap, or quadrupole/TOF instruments, fragment ion spectra are generated by a process called collision-induced dissociation (CID) in which the peptide ion to be analyzed is isolated and fragmented in a collision cell, and the fragment ion spectrum is recorded. Typically, but not exclusively, these types of mass spectrometers are used in conjunction with ESI. For the most part, the low-energy CID spectra of peptides generated by ESI-MS/MS are of high quality and are sequence specific. Other mass spectrometric methods including fragmentation of high-energy ions and post-source decay (PSD) in a MALDI MS also produce sequence-specific fragment ion spectra.[128] Generally, these are more difficult to interpret than the low-energy CID spectra generated by ESI-MS/MS and are not further discussed in this paper.

a. Background to Peptide Fragmentation. Tandem mass spectra generated by the fragmentation of peptide ions in the gas phase at low collision energy are dominated by fragment ions resulting from cleavage at the amide bonds. Very little amino acid side chain fragmentation is observed. Such spectra are much less complex than the high collision energy spectra generated in magnetic sector or TOF/TOF instruments. The low-energy CID spectra generated by the types of mass spectrometers most frequently used in proteomics are therefore relatively simple to interpret, and a straightforward nomenclature for annotating the MS spectra has been adapted (Figure 2). The nomenclature differentiates fragment ions according to the amide bond that fragments and the end of the peptide that retains a charge after fragmentation.[129,130] If the positive charge

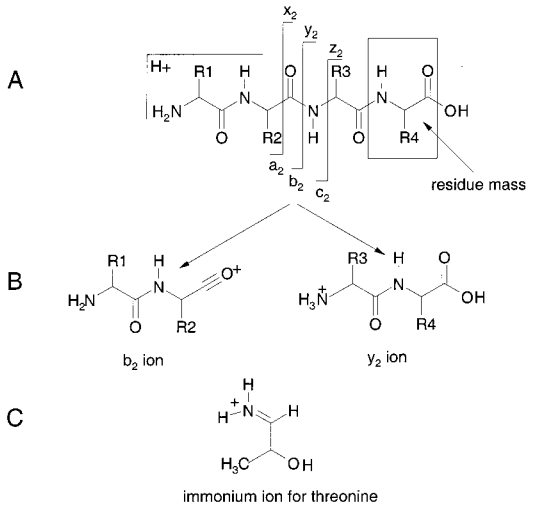


Figure 2. Peptide fragment ion nomenclature. (A) Nomenclature for peptide fragment ions that form via cleavage of bonds along the peptide backbone. (B) Example structure for *b* and *y* ions. Note that the b_2 ion specifically is thought to take the form of a cyclic oxazole[131] rather than a highly unstable acyl cation as shown. (C) Immonium ion for threonine drawn for simplicity. Not all amino acids generate immonium ions and then not to the same extent.

associated with the parent peptide ion remains on the amino-terminal side of the fragmented amide bond, then this fragment ion is referred to as a *b* ion. However, the fragment ion is referred to as a *y* ion if the charge remains on the carboxyl-terminal side of the broken amide bond. Since in principle every peptide bond can fragment to generate a *b* or *y* ion, respectively, subscripts are used to designate the specific amide bond that was fragmented to generate the observed fragment ions. *b* ions are designated by a subscript that reflects the number of amino acid residues present on the fragment ion counted from the amino-terminus, whereas the subscript of *y* ions indicates the number of amino acids present, counting from the carboxyl-terminus. These individual fragment ion *m*/*z* values as shown in Figure 1C can be easily calculated from the amino acid sequence, using the nominal (i.e., monoisotopic value rounded to an integer value) residue masses found in Table 2. To calculate the masses of the *b* ion series (Figure 1C), 1 u (for 1 H) is added to the nominal mass for the first residue. In the example indicated, the nominal mass for glycine (nominal mass = 58) is added to indicate the mass of the b_1 ion. To calculate the masses for the b_2, b_3, and following fragment ions, this process is continued by the addition of the nominal mass for the second, third, and following amino acid residues, respectively, until the final, carboxy-terminal amino acid is included. The *b* ion series will stop at 902 or 18 u short of the $[M + H]^+$ mass. To calculate the masses for the *y* ion series (Figure 1C), 19 u (for H_3O^+) is added to the nominal residue of the carboxy-terminal amino acid. In the example indicated, this residue is lysine with a mass of 147 u. As for the *b* ion series, this process is continued with the addition of the nominal mass of the following amino acids until the $[M + H]^+$ value is reached. While it is relatively simple to calculate the elements of the *b* and *y* ion series from the peptide sequence, it is much less straightforward to read the amino acid sequence from the CID spectrum of a peptide ion. This is mainly because peptide fragmentation under the conditions encountered in the collision cell of a mass spectrometer are sequence dependent, and the rules for fragmentation are not completely understood. Additionally, the drawing of the *b* ion structure as an acyl cation as presented in Figure 2 is historical. It has been recently shown that the b_2 ion specifically is much more likely to exist as a cyclic oxazole ring[131] rather than the acyl cation[54] as shown.

Table 2. Residue and Immonium Ion Masses of 20 Common Amino Acids

amino acid (3/1 letter codes)	nominal residue mass	immonium ion mass
alanine (Ala /A)	71	44
arginine (Arg/R)	156	129
aspartic acid (Asp/ D	115	87
asparagine (Asn/N)	114	88
cysteine (Cys/C)	103	76
glutamic acid (Glu/E)	129	102
glutamine (Gln/Q)	128	101
glycine (Gly/G)	57	30
histidine (His/H)	137	110
isoleucine (Ile/I)	113	86
leucine (Leu/L)	113	86
lysine (Lys/K)	128	101
methionine (Met/M)	131	104
phenylalanine (Phe/F)	147	120
proline (Pro/P)	97	70
serine (Ser/S)	87	60
threonine (Thr/T)	101	74
tryptophan (Trp/W)	186	159
tyrosine (Tyr/Y)	163	136
valine (Val/V)	99	72

b. Properties of Peptide Fragment Ion Spectra. The CID spectrum of a peptide ion acquired at low collision energy can be considered a composite of many discrete fragmentation events. Each peptide tandem mass spectrum will contain *b* and *y* ions as well as other fragment ions that can be used to interpret the amino acid sequence. These include diagnostic ions generated by the neutral loss of specific groups from amino acid side chains (e.g., the loss of ammonia (−17 u) from Gln, Lys, and Arg or of water (−18 u) from Ser, Thr, Asp and Glu) and low mass ions that result from the fragmentation of amino acids down to a basic unit consisting of the side chain residue and an immonium functionality (Figure 2). The *b* ion series also often shows a satellite ion series in which each signal is 28 u lower than the corresponding *b* ion. These signals result from the neutral loss of carbon monoxide and are referred to as an *a* ion series. CID spectra can be further complicated by the presence of internal fragment ions that represent some contiguous sequence of amino acids in the peptide. These are generated if a specific peptide ion undergoes two or more fragmentation events. Empirical observation shows that internal fragments often occur if either proline[132,133] or aspartic acid[134] residues are present in a sequence and even more so at any aspartyl-proline bond,[135] indicating that not all peptide bonds have the same propensity to fragment during low-energy CID. For the same reason, the relative

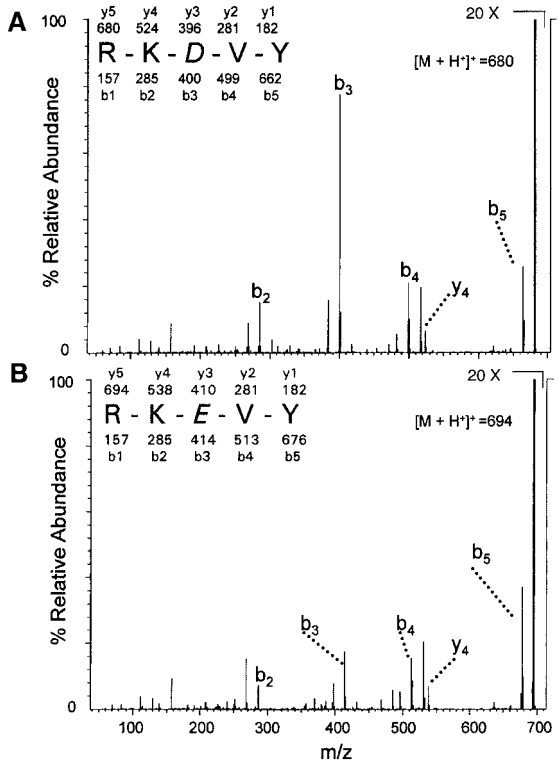


Figure 3. Tendency of peptides to fragment at aspartic acid. The $[M + H]^+$ ions for peptides, identical in sequence except for a substitution of aspartic acid (A) with glutamic acid (B), were subjected to CID under identical conditions using electrospray ionization on a triple quadrupole mass spectrometer.

intensity of fragment ions in peptide CID spectra is uneven and somewhat unpredictable. Some of the rules that control peptide ion fragmentation in a collision cell have been determined;[54,131,136] many others remain to be studied. If a proline residue is present in a peptide sequence, the most intense ions in the CID spectrum will generally be due to fragmentation on the amino-terminal side of proline.[132] This is thought to occur because the gas-phase basicity of the proline imide bond is greater than that for any of the amide bonds. Under the moving proton hypothesis for CID, the proton available for fragmentation is therefore statistically more likely to be at this imide bond as compared to the amide bonds in the peptide.[136] Additionally, it is known that peptides that contain aspartic acid tend to fragment at the carboxyl-terminal adjacent amide bond (i.e., Asp-Xxx; Figure 3A). This observation may be due to the ability of the aspartic acid side chain to influence the gas-phase basicity of the adjacent carboxyl-terminal amide bond via formation of a transient six-membered ring between the carboxylic acid group of the Asp side chain and the nitrogen of the adjacent amide bond. Leading credence to this is the observation that when glutamic acid, chemically similar to aspartic acid in that it contains one extra methylene carbon giving the side chain more degrees of freedom, is substituted for aspartic acid into a peptide, the Glu-Xxx bond (Figure 3B) does not fragment as readily as the Asp-Xxx bond (Figure 3A).[134]

Thus, the quality of peptide tandem mass spectra is dependent on the sequence location of amino acids, amino acid side chain basicity, amino acid side chain structure, and charge state of the peptide ion fragmented. If proteins are completely digested with trypsin, then lysine or arginine residues will be present at the carboxyl-terminus of all peptides except for the C-terminal peptide of the original protein. A charge sequestered by lysine or arginine at the C-terminus tends to produce a more complete series of *y* ion fragments than will be generated by peptides produced by protein digestion with chymotrypsin or other protease where lysine and arginine are distributed throughout the sequences rather than at the C-terminus. Additionally, $[M + 2H]^{2+}$ ions of peptides will produce tandem mass spectra of higher quality than those from either $[M + H]^+$ or $[M + 3H]^{3+}$ peptide ions. The $[M + 2H]^{2+}$ peptide ions fragmented under low-energy CID produce spectra, although there are exceptions such as when proline and/or histidine are internal to the peptide sequence, that contain $[M + H]^+$ fragment ions that are more readily interpreted than tandem mass spectra of $[M + 3H]^{3+}$ and higher charge states that produce multiply charged fragment ions.

c. Generation of Tandem Mass Spectra. Peptide fragmentation for the purpose of protein identification, either for single isolated proteins or on a proteome wide scale, is most often carried out by CID in a triple quadrupole (TQ),[47] ion-trap (IT)[58] or quadrupole time-of-flight (QTOF)[69,70] mass spectrometer and to a lesser extent by PSD-MALDI-MS.[128,137] Among these methods, fragmentation by PSD-MALDI-MS is least well-controlled partly because only a few parameters of the experiment can be readily varied (i.e., laser energy and type of matrix). Furthermore, peptide ionization by MALDI generates mostly $[M + H]^+$ ions that do not produce readily interpretable tandem mass spectra. PSD-MALDI-MS therefore has a lower success rate, and the spectra are in many cases of lower quality than the spectra of the same peptides produced by ESI and low-energy CID in a collision cell. For these reasons, CID of peptides in TQ, IT, and QTOF mass spectrometers is more frequently used for protein identification.

If one or a few peptides derived from a pure protein are being analyzed by CID, the time required for the mass spectrometer to select a specific peptide ion for CID, to fragment the ion, and to record the fragment ion spectrum is fast in comparison with the time required to prepare the sample and to analyze the data, even if computer algorithms are used for data interpretation. The identification of one or a few proteins is therefore frequently carried out in a manual mode in which the *m*/*z* ratio of the peptide ion selected for CID is controlled by the operator of the instrument. Manual precursor ion selection and control of CID conditions has the advantage that the fragmentation conditions can be optimized by an experienced operator for each precursor ion during the experiment. A particularly successful implementation of this approach uses a variation in ESI called nanospray in which a peptide sample is introduced at very low flow rates, typically nanoliters per minute, into the mass spectrometer.[53,138−140] The low sample consumption afforded by the nanospray tech-

nique allows for extended observation and accumulation of the ion signals and generally yields CID spectra of excellent quality.

Proteomics requires the analysis of large numbers of proteins, each one potentially generating multiple peptide fragments. The time required for the analysis of a single peptide by CID therefore rapidly becomes limiting in proteome studies if each ion has to be manually identified and selected. Therefore, protocols for automated, instrument-controlled selection of precursor ions have been developed. In these methods, ion selection for CID is under computer control and based on signals observed in the full-scan mass spectrum (i.e., data-dependent MS/MS).[58,141−145] In the most basic implementation of the method, the system selects the most intense ion (i.e., base peak) in a given *m*/*z* range for CID, carries out CID on that ion, and then writes that parent ion *m*/*z* value to a list for dynamic exclusion from further CID for some defined time. This iteration begins again as the next most intense ion in the original mass spectrum is chosen for CID. If the sample is introduced by infusion of an unseparated peptide mixture from a syringe or by nanospray, the system will walk through all ions above a preset threshold. If the sample is introduced from an on-line separation method such as capillary electrophoresis or HPLC, the observed mass spectrum will change continuously during the separation. If complex peptide mixtures are separated and analyzed on-line by ESI-MS/MS, it is frequently the case that a larger number of peptide ions are detected in a chromatographic peak than can be subjected to CID during the time available. Therefore, even with dynamic exclusion to limit redundant CID of the most abundant peptides, not all of the peptide ions with a signal intensity above the preset selection threshold will be analyzed. The consequence of this situation is an apparent compression of the dynamic range of the mass spectrometer. Some of the more sophisticated systems alleviate this problem by modulating the flow rate of the HPLC [58,142] or capillary electrophoresis[141] separation system into the ESI source in a signal-dependent manner. Generally, the flow rate is reduced as long as ion signals exceeding the preset threshold are detected in a chromatographic peak and re-accelerated between peaks. These "peak parking" procedures take advantage of the concentration-dependent nature of ESI[146] and allow tandem MS spectra to be acquired on some of the lower abundance species that might normally be missed.

For peptide mass mapping, the information collectively contained in the masses of several peptides derived from the same protein is used for protein identification by database searching. In contrast, the CID spectrum of a single peptide can, in principle, contain a sufficient amount of information for unambiguous identification of a protein.[147,148] Therefore, if a mixture of several proteins is concurrently digested, the components of the mixture can be identified based on the CID spectra, provided that at least one CID spectrum per protein is generated. It is therefore no longer necessary to separate proteins to homogeneity prior to proteolysis. The composition of even relatively complex protein mixtures can now be ascertained without purification of individual proteins by using two-dimensional chromatography methods coupled to data-dependent ESI-MS/MS.[127]

While most of the discussion has been about low-energy CID, it is worth noting a new type of sequencing mass spectrometer that provides CID at high energy using a TOF/TOF mass spectrometer.[73] This was achieved by coupling two TOF mass spectrometers together via a collision cell between them. The new design combines the advantages of MALDI such as high sensitivity for peptide analysis, relative insensitivity to salts, surfactants, and other contaminants with high-energy CID where amino acids such as isoleucine and leucine can be distinguished by side-chain fragmentation. As with other types of sequencing mass spectrometers, a complete CID spectrum can be acquired in a single acquisition, obviating the need to sum as many as 10 spectra as is necessary with PSD on a single TOF mass spectrometer. Additionally, the MALDI-TOF/TOF mass spectrometer promises to be capable of acquiring tandem mass spectra at a rate that is an order of magnitude above the capabilities of IT and QTOF instruments, which will be significant for proteome studies of mammals where the number of ORFs is 10−100 times that of microbials.

d. Protein Identification Using Tandem Mass Spectra. Regardless of the method of fragmentation, low-energy spectra will contain redundant pieces of information such as overlapping *b* and *y* series ions, multiple internal ions from the same peptide, and immonium ions. This redundancy makes fragment ion spectra an extremely rich source of sequence-specific information, but it also complicates the interpretation of the sequence. This is illustrated by the CID spectrum shown in Figure 4A that represents a low-energy spectrum of the $[M + 2H]^{2+}$ ion of angiotensin. Even with the peptide sequence available to aid spectral interpretation, there are only a few ions that can be easily assigned unambiguously without considering the possibility of doubly charged fragment ions. Furthermore, it is not immediately apparent to which ion series a particular ion belongs (i.e., is it a *y* or *b* ion?); that is to say that directionality is not apparent a priori from the fragment ions. While the *b* ion series will often be accompanied by an *a* ion series, for a complete unknown even this clue may not be of much help because not every *b* ion will have an associated *a* ion. In the case illustrated, the abundance of doubly charged ions arise from the amino acids arginine, proline, and histidine that can serve as sites for proton sequestration. An additional complication arises from the fact that this peptide was not produced by cleavage of a protein with trypsin. Therefore, the peptide does not contain a carboxyl-terminal sequestered charge that helps reveal the *y* ion series and, conversely and even worse for interpretation, it contains an internal arginine. Compare the spectrum in Figure 4A to the one in Figure 4B, which is from the $[M + 2H]^{2+}$ ion of peptide resulting from autolysis of trypsin. This tryptic autocatalytic peptide has no charged residues

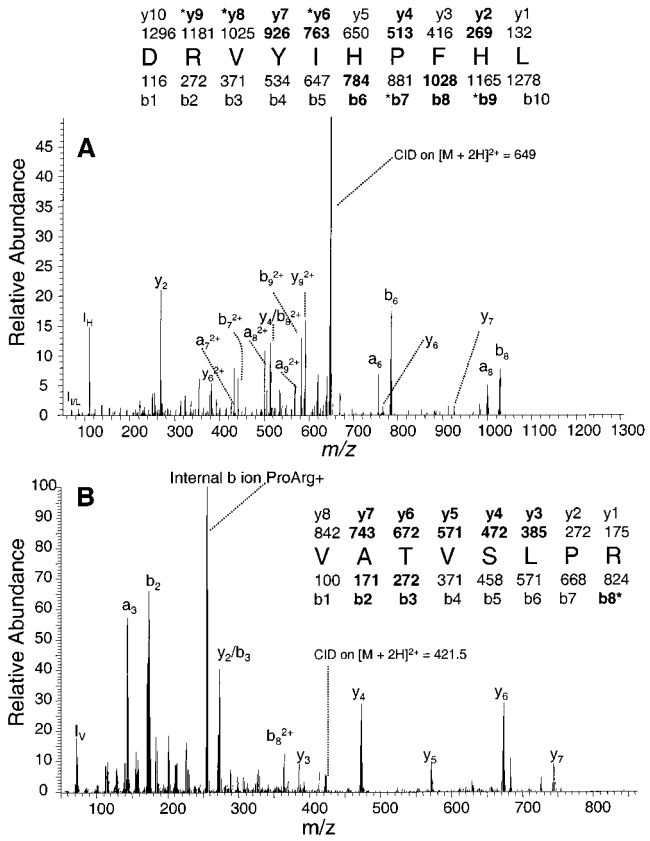


Figure 4. Annotated tandem mass spectra**.** The annotated tandem mass spectra for (A) $[M + 2H]^{2+}$ of angiotensin and (B) a peptide from the protease trypsin commonly produced by autocatalysis of trypsin. Both spectra were obtained at low collision energy using electrospray ionization on a triple quadrupole mass spectrometer.

located in the interior of the sequence and has arginine at the carboxyl-terminus. Both of these factors are advantages that help one to distinguish the *y* ion series over the *b* ion series.

From this example, it is apparent that the manual, explicit interpretation of sequence information contained in a CID spectrum is complex, labor- and decision-intensive, and slow. Clearly, the manual interpretation of CID spectra would constitute a significant bottleneck in every proteome project based on protein identification by CID spectra. Since at least some of the rules that determine the fragmentation of peptide ions in a low-energy collision cell have been known, it seemed sensible to devise computer algorithms that interpret or at least assist an operator in the interpretation of the CID spectra. While such attempts at computerized de novo sequence interpretation have been progressing, the most significant advances in rapid protein identification using CID spectra came from the development of algorithms that correlate peptide CID spectra with sequence databases, automatically and without the need for user input.

2. De Novo Peptide Sequencing

The basic problem presented by the de novo interpretation of a CID spectrum of a peptide with unknown sequence is one of directionality. It is a priori impossible to determine which ions belong to the *b* ion and which ones belong to the *y* ion series; without this assignment, the sequence cannot be interpreted. Therefore, a number of methods have been developed that address this problem. They can be categorized into MS- and MS/MS-based methods. The MS-based methods attempt to derive sequence information without tandem MS. They are based on the generation of peptide ladders in which individual elements of the ladder differ in length by one amino acid. The mass differences between the elements of the ladder as determined by MS therefore indicate the amino acid sequence of the peptide analyzed. In the MS/MS-based approaches, CID spectra are generated as described above. To identify the *b* and *y* ion series, respectively, and therefore the directionality of the peptide sequence, a number of chemical modification procedures have been introduced.

Peptide ladders for sequence analysis without using tandem MS can be generated by chemical or enzymatic degradation of peptides and can proceed from either the amino- or the carboxyl-terminus. The chemical methods for determining amino-terminal sequences employ the principles of the Edman degradation chemistry and therefore require a free primary amine at the amino-terminus. Two variations of the method have been described. In the first, Edman reagents that react with the amino-terminal primary amine are added at the beginning of each sequencing cycle. A "ladder" of ions reflecting the peptide sequence is generated by blocking the amino-terminus of a small fraction of the peptides in each cycle via the addition of a small amount of isocyanate blocking agent to the Edman reagent. The blocked peptides are not degraded during subsequent cycles. After a number of degradation cycles, the sample is recovered and analyzed by MALDI-MS. The sample is a mixture of peptide fragments (i.e., a peptide ladder), the masses of which differ by the mass of the amino acid residue cleaved off in each sequential cycle.[149,150] The second approach, referred to as "nested peptide sequencing", is based on the addition of an aliquot of the peptide/protein at each cycle of the Edman degradation process. In this method, the Edman chemistry process is driven to completion using an excess of a volatile reagent that can be removed in each cycle by evaporation. Just as with ladder sequencing, a peptide sequence is generated when masses of products are measured.[151] Both methods require a free amino-terminus, and neither can resolve the isobaric residues isoleucine and leucine. However, glutamine (residue mass 128.13) and lysine (residue mass 128.17), difficult to distinguish in CID spectra, can be easily distinguished because the ϵ-amino group of lysine side is modified with the Edman reagent. In principle, both methods can cope with mixtures of peptides whose intact masses are sufficiently different to avoid overlap of the degradation products.[150] Alternatives to the chemical stepwise degradation peptide sequence ladders have also been generated by truncation of peptides by amino and carboxyl peptidases. This approach has the advantage that reactions can be conducted with very small quantities of starting material and can be carried out directly on the

MALDI probe surface. The enzymatic reactions are stopped by the addition of a matrix,[115−117] and the method is easily amenable to time course studies. Neither the chemical nor the enzymatic methods to generate sequence ladders for peptide sequencing by MS are routinely used.

If high-quality CID spectra can be obtained by MS/MS (i.e., spectra containing complete *b* and *y* ion series), it may be possible in some cases to determine de novo the peptide sequence.[152] Unfortunately, interpretation of CID spectra is often a difficult process even with high-quality data. Confidence in the final sequence assignment may remain low unless a biological assay is available to test the proposed sequence. It is much more common for CID to be of marginal quality due to poor ion statistics arising from either too little peptide or a peptide chemistry that is partially refractory to fragmentation. The difficulty of de novo sequence analysis, and in particular of identifying the respective ion series, has led to several ingenious methods that help elucidate the sequence. For instance, fragment ions belonging to a *y* ion series can be identified by conducting proteolysis with trypsin in a buffer containing 50% v/v $H_2{}^{18}O$/50% v/v $H_2{}^{16}O$. This takes advantage of the hydrolytic action of trypsin that adds a molecule of water across each amide bond that is hydrolyzed. Every peptide with the exception of the peptide derived from the carboxyl-terminus will appear as doublets differing in mass by 2 u.[153−155] If both isotopically labeled parent ions are selected for CID together, only fragment ions with an intact carboxyl-terminus will appear as doublets separated by 2 u. While the method simplifies the fragment ion assignment, it also necessarily reduces the signal intensity of each *y* ion, therefore making sequencing at very high sensitivities more difficult. The use of high-resolution MS instruments such as the QTOF tandem mass spectrometer for analysis of isotopically labeled *y* ions as described above can result in complete interpretation of peptide spectra in a single experiment.[156] The method has also been successfully applied with an ion-trap mass spectrometer.[157] An alternative chemical approach that again distinguishes *y* ions from *b* ions involves methyl esterification of the carboxyl groups in a peptide. This reaction increases the mass of the peptide by 14 u for each carboxyl group (unmodified carboxyl-terminus, side chains of aspartic acid, glutamic acid).[54] If there are no acidic residues in the peptide and only the carboxyl-terminus is esterified, the mass of each signal in the *y* series will be increased by 14 u as compared to the corresponding signals obtained from the unmodified peptide. To distinguish the *y* series ions, this approach therefore requires a comparison of the fragment ion spectra obtained from the methylated peptide and the original, underivatized peptide. Methods to specifically tag the amino-terminal residue with the purpose of identifying the *b* ion series have also been utilized. In these experiments, peptide amino groups were derivatized with a reagent containing a permanent positive charge in the gas.[54,111,158] Additionally, a method that takes advantage of the moving proton hypothesis for fragmentation of tryptic peptides was described that derivatizes the amino-terminus with an acidic reagent.[159] This approach helps achieve a charge balance between the basic carboxyl-terminal residues of either lysine or arginine (in tryptic peptides) and the moving proton available for fragmentation.

With the development of computer algorithms and large-scale databases, the need for de novo sequencing of peptides is clearly declining. However, for the sequence analysis of the many proteins from species for which no genomic or expressed sequence tag database is available, the art of reading out amino acid sequences from CID spectra will remain important. An excellent, detailed tutorial was recently published.[160]

3. Manual Generation of Peptide Sequence Tags

In an attempt to accelerate protein identification using CID spectra of peptides and to take advantage of the extensive sequence databases available, two basic types of approaches have been employed. They have in common that they correlate the information in the fragment ion spectra with sequence databases using computer algorithms. They differ in the way the information is extracted from the CID spectra. The first relies on a partial manual interpretation of the spectrum to identify consecutive elements of a particular (*b* or *y*) ion series to provide a partial sequence (i.e., any contiguous set of *b* or *y* ions). This partial sequence is then used together with the mass of the parent ion mass to determine by subtraction the mass difference between the parent ion mass and the total mass of the amino acids that constitute the partial sequence tag. This calculation provides a mass the sum of which constitutes a possible amino acid composition. If the protein was digested with trypsin, then the amino acid in the carboxyl-terminal position can be guessed as either arginine or lysine. Together all of this information provides a peptide-sequence tag that can be used for searching databases.[161,162] Using the example in Figure 1C, a partial internal sequence tag, His-Glu, may be derived from observing three *y* ions at *m*/*z* 726, 589, and 460. This along with the parent peptide mass, 920 *m*/*z*, provides residue masses with unknown sequence at the amino- and carboxyl-terminus of 194 and 441, respectively. This information is then used as input for an algorithm to search a database for protein identification. More recently, the generation of sequence tags has been automated, further increasing the utility of this approach.[163] A recently published algorithm, SHERENGA, intended to automate the interpretation of CID spectra for de novo sequencing currently falls short of this ambitious goal but is very useful for the generation of sequence tags or for validating sequence database matches generated by automated database searching tools (see below).[164] SHERENGA automatically learns fragment ion types (i.e., *b* vs *y* ion directionality) and intensity thresholds from a database of spectra generated from peptides of known sequence. It can be applied to data from any type of mass spectrometer provided that the spectra in the database are generated by the same type of instrument as the spectra to be analyzed.

4. Automated Interpretation of CID Spectra

The second approach for computer-assisted protein identification via CID spectra uses the uninterpreted fragment ion pattern and mass of the parent ion as input for sequence database searching. This approach is exemplified by the algorithm SEQUEST.[165] The algorithm first creates a list of peptide masses isobaric to the observed mass on which CID was carried out by searching the database of choice for possible amino acid sequences that can generate peptide masses to match the mass of the parent peptide. For each of these candidate peptides, the program calculates the masses of the fragment ions expected, without consideration of chemical information that is reflected in the relative intensities of fragment ions (i.e., all predicted fragment ions are equal in intensity) and generates a theoretical CID mass spectrum for comparison. The program then compares the observed fragment ion spectrum with the top 500 theoretical fragment ion spectra using cross-correlation algorithms. Each comparison then receives a score that is ranked relative to all other possibilities according to a number of parameters such as the number of fragment ions predicted versus found. Once an answer is arrived at, then it is important to confirm this top score using a functional assay if possible and, at the very least, to manually check the predicted sequence.[147,166] The constraints on database searching of a given stretch of peptide sequence are so powerful that the tandem MS spectrum of a single peptide can be adequate for protein identification in an EST database.[147] The approach is easily automated[167] and can also be adapted to find peptides carrying specified post-translational modifications by instructing the program to anticipate modification at specific residues (e.g., 80 u is added to phosphorylated residues such as serine, threonine, or tyrosine).[167,168] A list of some Internet sites with protein identification resources developed by these and other investigators can be found in Table 1.

5. Accurate Mass Tags

A type of instrument that is gaining in popularity for proteomic analysis is FT-ICR-MS. The advantages of FT-ICR-MS are severalfold and include high resolution over a broad m/z range,[74] high sensitivity/dynamic range,[62] and high mass accuracy.[109,110,169] Both MALDI and ESI-FT-ICR MS are proving to be an order of magnitude more sensitive for peptide detection than standard triple quadrupole and ion-trap technology. As little as tens of attomoles of peptide loaded onto a MALDI probe can be detected.[170] Additionally for ESI, 10 amol of phosphoangiotensin II loaded on a 50 μm i.d. capillary column and eluted via a standard acetonitrile gradient was used to measure masses and fragment peptides via ESI-FT-MS.[110,171] The ability to measure peptide masses to 0.1 ppm at 1000 u creates the possibility of using the mass of a single peptide as a unique identifier for a protein when working with a specific genome of relatively small size such as *S. cerevisiae* (~6000 ORFs) or *H. influenza* (~1700 ORFs). Whereas the peptide mass normally only provides a subset of possible amino acid compositions[122] that could be combined to define the measured mass, very high mass accuracy allows the use of accurate mass tags (AMTs) for protein identification in small genomes but probably will not be adequate for larger mammalian genomes such as humans, which may have anywhere from 40 000 to 100 000 ORFs (as of 2000, the number of human ORFs is disputed).[109] Furthermore, in cases where the mass measured to an accuracy of 0.1 ppm is not unique, it may be possible to use readily available constraints such as the presence of cysteine[110] in a peptide to positively identify the parent protein. The advantage of protein identification from the accurate mass of single peptides circumvents the most significant disadvantage of data-dependent MS/MS, which is that it is impossible (no matter how many tricks are tried) to perform tandem MS on every ion presented in a chromatographic window in time.[58,141,143] Additionally, while a specific peptide ion is selected for CID, other coeluting ions are not selected for tandem MS and are therefore excluded for analysis. These excluded ions may or may not be selected for tandem MS in subsequent iterations of the process. By circumventing the need for data-dependent tandem MS, all ions present in a chromatographic window in time are used for protein identification. This means that low abundance proteins that are passed over by the data-dependent methods because of low signal intensity can be identified by taking an accurate mass snapshot of each chromatographic window.[110]

C. Protein Identification in Complex Mixtures

The field of proteomics has grown out of the mature technology of high-resolution two-dimensional gel electrophoresis (2DE) for protein separation and quantitation[172,173] and the increasingly refined technologies described above for the identification of separated proteins.[85] Today, 2DE and protein MS represent an integrated technology by which several thousand protein species can be separated, detected, and quantified in a single operation, and hundreds of the detected proteins can be identified in a highly automated fashion by sequential analysis of the peptide mixtures generated by digestion of individual gel spots. It is commonly assumed that 2DE-MS is a suitable technology base for global proteome analysis based on its ability to display, quantify, and identify thousands of proteins in a single gel.[91] However, closer examination of the proteins routinely identified by proteome studies suggests that 2DE-MS does not represent a truly global technique. Specific classes of proteins are known to be absent or underrepresented in 2D gel patterns. These include very acidic or basic proteins, excessively large or small proteins, and membrane proteins. In addition, by examining codon bias values of proteins identified from 2D gels, it has now been shown that the 2DE-MS approach is incapable of detecting low abundance proteins without pre-gel enrichment.[174] Codon bias is a measure of the propensity of an organism to selectively utilize certain codons, which result in the incorporation of the same amino acid residue in a growing

polypeptide chain. It is also thought to be a good measure of protein abundance because highly expressed proteins generally have large codon bias values.[175] It has been previously shown that almost without exception proteins identified from 2DE-MS experiments of whole yeast lysates are abundant proteins (codon bias values >0.2).[91,174,176,177] This is a striking finding since more than one-half of all yeast genes have codon bias values <0.1, thus making these proteins undetectable by 2DE without prior enrichment. Furthermore, accurate quantitation of proteins separated by 2DE is complex and of limited dynamic range, particularly if high-sensitivity staining methods are being used.[178] Clearly, the detection and quantification of low abundance proteins such as trnscription factors, protein kinases, and other proteins of regulatory function is an important component of proteomics and incompatible with the standard 2DE-MS approach.

These limitations inherent in the 2DE-MS or 2DE-MS/MS approach to proteomics and the emerging ability to identify the components in protein mixtures using data-dependent, automated LC-MS/MS and sequence database searching have catalyzed the development of new, chromatography-based methods for the identification of the proteins contained in complex mixtures without the need for separation of the mixture into the individual protein components.[127,144,179−181] This is accomplished by the digestion of the unseparated proteins and the analysis of the resulting complex peptide mixture by LC-MS/MS. What is most exciting about these types of experiments is the sheer numbers of peptides that can be sequenced automatically in reasonable time frame in a single analysis. Using an ion trap mass spectrometer, an MS strategy was reported for highly complex peptide mixtures that employs a single MS scan followed by five MS/MS (sequencing) scans on the five most-intense peptide ions in that scan. Up to 10^4 sequencing attempts were recorded in a single LC-MS analysis of 60-min duration.[143] To increase even further the number of peptides that can be sequenced in a single analysis, techniques where complex peptide mixtures are separated on-line by a combination of cation-exchange and reverse-phase chromatography have been reported.[127] The success of these methods to identify and catalog rapidly and reliably large numbers of proteins in complex mixtures makes them powerful tools for descriptive proteomics.

For proteomics in general, the separation sciences continue to make great strides in analyzing complex mixtures and offer the potential for circumventing gel electrophoresis as a preparative tool for MS. Over the past decade, gel electrophoresis followed by proteolysis of individual stained protein bands has been the most common method for separating proteins prior to MS identification.[182] However, a number of laboratories have been investigating the use of chromatography only based approaches that bypass the electrophoretic based preparative gel methods altogether, except for diagnostic purposes. Two-dimensional or orthogonal chromatography approaches such as cation exchange followed by reverse-phase on-line with tandem MS have been successfully used to identify proteins in complex mixtures after proteolysis.[127,183,184] Even more complex approaches have used computer-controlled setup with an autosampler, five columns, and three 10-port switching valves to allow a series of steps to be performed on-line, obviating the need for any manual transfers of materials. The strategy included the following: immunoaffinity chromatography, desalting and buffer exchange on a mixed-bed strong ion-exchange absorbent, enzymatic digestion on an immobilized trypsin column, capture of peptides on a short perfusion capillary reversed-phase column, and final separation on an analytical reversed-phase column with on-line MS/MS analysis.[185] These orthogonal chromatography techniques have as a common goal the circumvention of the weakness of data-dependent analysis of complex mixtures, namely, that very complex mixtures of peptides exceed the capacity of these computer routines to carry out CID on all of the peptides present in a given full-scan mass spectrum. Thus, by fractionating complex peptide mixtures on-line, these two-dimensional chromatography methods help extend the dynamic range of the overall analysis.

The reduction of the complexity of ions presented during any chromatographic window in time for the data-dependent MS routines that automatically select ions for fragmentation is an important focus in the further development of protein mixture analysis by LC-MS/MS or LC/LC-MS/MS. This is being approached by chemistry-based reduction of the peptide sample complexity and by the development of mass spectrometers that can more completely sample the peptides present in a sample. Selective tagging of specific, relatively rare functional groups in peptides has been introduced as a strategy to reduce the complexity of peptide samples. Reports to date used relatively specific alkylation reactions to selectively tag the sulfhydryl side chains in cysteine-containing peptides, but any other specific reaction targeting a rare group in a peptide would be equally suitable.[110,182,186] If MALDI-TOF/TOF instruments,[73] capable of conducting tandem MS at rates an order of magnitude faster than currently possible with IT or QTOF instruments, are combined with existing separation systems to fractionate peptides separated by HPLC or ion exchange directly onto a MALDI target, then more complete sequence coverage for any given sample should be one obvious result. This new design of sequencing mass spectrometer selects ions for CID with a TOF mass spectrometer and then analyzes fragment ions in a second TOF mass spectrometer rather than post-source decay.[187] Combining with pre-fractionation from HPLC separations would allow one fraction to be analyzed repeatedly and perhaps completely, thus obviating the need to repeat a chromatographic separation. Furthermore, it is possible that for proteomic studies where genomes are completely sequenced that developments in FT-ICR-MS may circumvent the need for "serial" tandem MS to identify proteins. It has recently been demonstrated that simultaneous or "parallel" trapping of a group of unrelated peptides (i.e., not derived from the same parent protein) followed by collective fragmen-

tation of all peptides in the ICR cell can be used to identify all proteins from which each peptide was derived without true tandem MS (i.e., selection of a single ion followed by fragmentation).[188] In this case, high mass accuracy provided a critical benefit in that it allowed simultaneous identification of multiple and different proteins using only a single peptide ion mass determined prior to fragmentation to 1 ppm together with a single peptide fragment ion mass also determined to 1 ppm. It is at least hypothetically possible that a similar approach could be implemented at lower mass accuracy on any TOF mass detector using in-source CID to collectively fragment peptides.

D. Analysis of Protein Expression

A promising and exciting new use for MS in proteomics involves not just the identification of proteins but also the determination of protein expression levels (relative quantity) between two different pools of proteins. Obtaining expression data for proteins as is routinely done for mRNA is important because protein expression levels often are diagnostic of a given cellular state and are not directly related to levels of mRNA expression.[174] Methods exist for the global comparison of mRNA as a function of cellular state,[189] and they are widely used to identify clusters of genes for which expression is idiotypic of a cellular state. Until recently, no such methods with adequate dynamic range existed for obtaining relative protein expression measurements for most or all the proteins expressed by a cell or tissue (see discussion related to dynamic range of 2DE). To add a quantitative dimension to non-2DE-based proteome analyses, the venerable technique of stable isotope labeling has been adapted for protein analysis.[190] The method involves the addition to a sample of chemically identical but stable isotope (e.g., ^{2}H, ^{13}C, ^{15}N, etc.) labeled internal standards. In the case of ESI and MALDI, ionization efficiency can be quite variable for peptides of different sequence or even identical peptides from different MALDI spots or for ESI different HPLC conditions. Thus, the best internal standard for a candidate peptide is a peptide of identical sequence but labeled with stable isotopes. Quantitative protein profiling is therefore accomplished when a protein mixture (reference sample) is compared to a second sample containing the same proteins at different abundances and labeled with heavy stable isotopes (Figure 5). In theory, all the peptides in the sample then exist in analyte pairs of identical sequence but different mass. As the peptide pairs have the same physicochemical properties, they are expected to behave identically during isolation, separation, and ionization. Thus, the ratio of intensities of the lower and higher mass components provides an accurate measure of the relative abundance of the peptides (and hence the protein) in the original protein mixtures. Three groups initially and independently reported measuring quantitative protein profiles based on stable isotopes.[186,191,192] The techniques differ in the method of incorporation of heavy isotopes (i.e., in vivo (Figure 5A) or in vitro (Figure 5B)) and in the analytical procedures used to measure protein expression and identify proteins. Chait and co-workers grew one yeast culture on medium containing the natural abundance of the isotopes of nitrogen (^{14}N, 99.6%; ^{15}N, 0.4%) while another culture was grown on the same medium enriched in ^{15}N (>96%).[191] After an appropriate growing period, the cell pools were combined, and proteins of interest were extracted and separated by RP-HPLC and then by SDS−PAGE. In-gel digestion of excised spots of interest resulted in peptide fragments, which were identified by peptide mass mapping. Each ^{15}N incorporated shifted the mass of any given peptide upward, leading to a paired peak for each peptide. The percent error of the experimental technique was found to be excellent (±10%). Smith and co-workers used stable isotope-depleted media to impart a specific isotope signature into proteins.[192] They compared the cadmium stress response in *Escherichia coli* grown in normal and rare isotope-depleted (^{13}C-, ^{15}N- and ^{2}H-depleted) media. Intact protein mass measurements were carried out by FT-ICR-MS. While no protein was positively identified, the expression ratios for 200 different proteins were compared. Clearly, stable isotope metabolic protein labeling using ^{15}N-enriched or depleted media permits quantitative protein profiling in conjunction with either

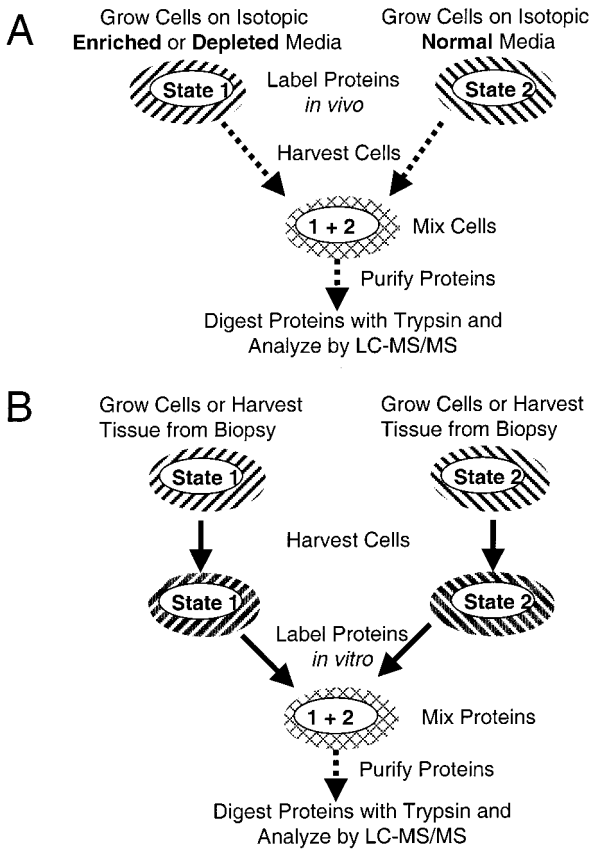


Figure 5. Scheme for determining protein expression. In the absence of a protein array on a chip as currently exists for measuring mRNA expression via cDNA arrays, MS as a universal detector can be used to determine relative protein expression between two samples of interest. (A) In vivo and (B) in vitro isotopic labeling of proteins followed by protein purification, proteolysis, and analysis by LC-MS/MS to determine both the identity of proteins and relative ratio of proteins expressed.

2DE or other separation techniques. However, this method has several disadvantages. First, the method does not allow for the analysis of proteins directly from tissue. Second, the stable isotope-enriched media are costly and may themselves affect cellular growth and protein production. Third, the increase in nominal mass due to stable isotope incorporation is not known until the sequence is determined. Therefore, protein identification has to precede quantitation. We have recently published a novel method for quantitative protein profiling based on isotope-coded affinity tags (ICAT).[186] In this method, the stable isotopes are incorporated post-isolation by selective alkylation of cysteines with either a heavy (d8) or normal (d0) reagent. The two protein mixtures are then mixed. At this point, any optional fractionation technique can be performed to enrich for low abundance proteins or to reduce the complexity of the mixture, while the relative quantities are strictly maintained. Prior to analysis, the protein mixture is digested with trypsin and passed over a monomeric avidin−agarose column. Because the ICAT label contains the stable isotope information as well as a biotin tag, ICAT-labeled (cysteine-containing) peptides are selectively isolated for analysis by microcapillary LC-ESI-MS/MS. The ratio of ion intensities from coeluting ICAT-labeled pairs permits the quantification while a subsequent MS/MS scan provides the protein identification. The advantages of the ICAT strategy are severalfold. First, the method is compatible with any amount of protein harvested from bodily fluids, cells, or tissues under any growth conditions. Second, the alkylation reaction is highly specific and occurs in the presence of salts, detergents, and stabilizers (e.g., SDS, urea, guanidine hydrochloride). Third, the complexity of the peptide mixture is reduced by isolating only cysteine-containing peptides. Fourth, the ICAT strategy permits almost any type of biochemical, immunological, or physical fractionation that makes it compatible with the analysis of low abundance proteins. There are two disadvantages to the method. First, the size of the ICAT label (~500 Da) is a relatively large modification that remains on each peptide throughout the MS analysis. This can complicate the database searching algorithms, especially for small peptides (<7 amino acids). Second, the method fails for proteins that contain no cysteines. However, only a small percentage of proteins are cysteine-free (8% in yeast), and ICAT reagents with specificities to groups other than thiols could be synthesized.

A third approach to determine protein expression levels that does not involve stable isotopic labeling is possible, but it requires that the sample number is high enough for statistical significance to be proven.[193,194] In this case, normalization between cellular states relies on proving that one MS signal is statistically higher or lower in ion current than the signal for the same peptide from a different sample. Obviously, these studies are inherently difficult to conduct given the "relative" nature of MS detectors. However, MS has the advantage of being a universal detector and will continue to play an important role in measuring protein expression until a protein chip or other device is designed that can fill this role. Thus, regardless of the method, quantitative analysis of proteomes promises to provide a complimentary technique to mRNA expression for developing clinical diagnostics and studying basic genetics.

III. Proteomes and Post-Translational Modifications

A. Proteomes

A proteome is neither a static entity nor the product of the direct translation of gene sequences into protein sequences. In the previous sections, we discussed MS-based methods for the identification of proteins in complex samples and for monitoring quantitative changes in their abundance. In this section, we will discuss the challenges posed by the diverse post-translational mechanisms that process and modify proteomes permanently or reversibly and current methods used for the analysis of the products of these mechanisms. MS is an essential component of virtually every current strategy but is by itself insufficient to analyze post-translational modifications and processing.

1. The Analytical Challenge

If all the relevant properties of proteins were apparent from the gene sequence and could therefore be precisely predicted, the justification for establishing complex and expensive platforms for proteome analysis would be low. Proteome analysis is based on the expectation that the information gained by direct protein analysis exceeds or complements that obtained by the more readily available methods for gene sequence analysis. In addition to the sequence and abundance, the properties of proteins that are of particular interest to biologists include their subcellular location and state of modification, their function and state of activity, and the nature of interacting proteins. It is not obvious how these diverse properties can be determined systematically and quantitatively for a single protein. Extending these measurements to a proteome wide scale is even more challenging and, despite recent advances, remains largely unachievable with current methods. In this context, applications of MS have mostly focused on the characterization of protein−protein complexes and the analysis of post-translational modifications. These two topics will be further discussed in this paper. It can also be anticipated that in the near future emerging technologies will be combined with MS to extend the range of proteome-wide analyses that indicate the functional state of a biological system. These include the ability to measure quantitatively the specific activity of specific classes of enzymes in complex protein samples[195] and methods such as laser capture microdissection[196,197] and advanced methods for subcellular fractionation and the isolation of cellular organelles[198] to determine the subcellular location of proteins and the dynamics of protein trafficking patterns. We expect that the development of proteomics technologies related to the

analysis of post-translational processing and control will be very exciting and fruitful in the next few years.

2. Analysis of Protein–Protein Complexes

Most cellular functions are not performed by individual proteins but rather by protein assemblies also termed multi-protein complexes. It is therefore frequently assumed that proteins that specifically interact also partake in the same function, and the identification of specifically interacting proteins is an important component of the proteomics quest to study the function of biological processes. In general, the methods described above for the analysis of protein mixtures in general are also suited for the analysis of protein complexes, and some of the most scientifically rewarding applications of protein MS have involved the analysis of protein complexes. Yates and co-workers[127] identified more than 70 proteins present in the yeast ribosome in a single analysis using LC/LC-MS/MS. Peptide mass mapping was used to exhaustively analyzed the composition, architecture, and transport mechanism of the yeast nuclear pore complex.[199] Chemical cross-linking and MS were used to examine the spatial organization of multi-protein complexes,[44] and the components of the T-cell receptor complex[200] have been study by SDS−PAGE and tandem mass spectrometry. Such projects critically depend on the ability to isolate the target complex cleanly and in good yields. To this end, a tandem affinity purification (TAP) method has been developed and demonstrated its effectiveness by examining the yeast spliceosome.[201,202] In eukaryotes, seven Sm proteins bind to the U1, U2, U4, and U5 spliceosomal snRNAs while seven Sm-like proteins (Lsm2p−Lsm8p) are associated with U6 snRNA. Another yeast Sm-like protein, Lsm1p, does not interact with U6 snRNA. Using the tandem affinity purification (TAP) method, Lsm1p was identified among the subunits associated with Lsm3p. Coprecipitation, using antibody (so-called immunoprecipitation) experiments, demonstrated that Lsm1p together with Lsm2p−Lsm7p formed a new seven-subunit complex. The two related Sm-like protein complexes were purified, and the proteins recovered were identified by MS. MS and the TAP purification scheme were thus used to confirm the association of the Lsm2p-Lsm8p complex with U6 snRNA. Approaches such as this can be used in conjunction with and to test results from purely genetic methods such as the yeast two-hybrid approach.[203] Coprecipitation methods combined with mass spectrometric identification of proteins complement the genetic approaches such as the TAP purification scheme and the popular yeast two-hybrid system, which can be set up in a high throughput format but suffers from a lack of specificity. The above discussion of just one simple pathway reveals the complexity involved in trying to understand protein−protein interaction pathways on a global scale. In the coming years, proteome laboratories will attempt to map out all known pathways in select cellular systems using MS, genetics, and molecular biology.

B. Post-Translational Modifications

1. Background

MS is the method of choice for the detection and identification of post-translational modifications (PTMs). In principle, the methods used for protein identification are also applicable to the analysis of PTMs. For a number of reasons, PTM analysis is however significantly more complex than simple protein identification: (i) Proteins are frequently modified to a low stoichiometry only. Therefore, a high sensitivity of detection for the modified peptides is required. (ii) While proteins can be identified by the sequence or the CID spectrum of a single peptide, the identification of PTMs requires the isolation and analysis of the specific peptide that contains the modified residue(s). (iii) The bond between the PTM and the peptide is frequently labile. It may therefore be difficult to find conditions that maintain the peptide in its modified state during sample work up and ionization. (iv) More than 200 different types of protein modifications have been described.[204] The total sequence space containing all the potential modified protein sequences is therefore enormous. Likewise, the space required to comprehensively treat the procedures for the analysis of all types of post-translational modifications by far exceeds the space available for this chapter. We focus this discussion on protein phosphorylation because it is both biologically important and illustrates the challenges posed by PTM analysis. Many of the general approaches and specific methods discussed in the context of protein phosphorylation are however directly or with minor adaptations applicable to the analysis of different types of modifications.

a. Protein Phosphorylation: Catalysis and Biology. Among the large number of PTMs described to date, only a few have been shown to be reversible and thus potentially of regulatory importance in biological processes. Of these, protein phosphorylation has received the most attention and is the best understood with respect to both the enzymes involved catalyzing the phosphorylation/dephosphorylation reactions and the functional consequences of protein phosphorylation. The most common type of protein phosphorylation studied involves the formation of phosphate esters bonds with the hydroxyl side chains of serine, threonine, and tyrosine. Two counteracting enzyme systems, kinases and phosphatases, catalyze protein phosphorylation and dephosphorylation, respectively. The structures, specificities, and regulation of the most common of these is well-studied and reviewed.[205−207] There are assumed to be hundreds of protein kinases/phosphatases differing in their substrate specificities, kinetic properties, tissue distribution, and association with regulatory pathways. For example, analysis of the complete genomic DNA sequence from *S. cerevisiae* for sequence motifs that are thought to be indicative of protein kinases (and phosphatases) predicts 123 different protein kinases (and 40 protein phosphatases) that could be expressed. Thus, approximately 2% of expressed yeast proteins are involved in protein phosphorylation reactions, and presumably a much larger number of

proteins are phosphorylated under specific physiological conditions. In addition to phosphate esters involving the side chains of the hydroxyl amino acids, phosphoramidates of arginine, histidine, and lysine and acyl derivatives of aspartic and glutamic acid have also been observed. Some of these modifications are not typically observed unless specific precautions are taken to prevent their elimination during protein isolation.[208] For example, phosphohistidine is a relatively common modification, at least in prokaryotes, but it is completely eliminated by the acidic conditions commonly used for protein staining in polyacrylamide gels.

Protein phosphorylation generally exerts it regulatory function by altering the structure and thus the function of target proteins.[207,208−210] The functional consequences of site-specific protein phosphorylation are dramatic and diverse. Glycogen phosphorylase is the prototypical enzyme that is controlled by phosphorylation.[206] A single phosphorylation event involving a serine residue close to the N-terminal induces a long-range allosteric change and thus a change in the catalytic activity.[211] Work on the protein tyrosine kinase p56lck demonstrated that phosphorylation at a single serine, possibly by a mitogen-activated protein (MAP) kinase,[212] can alter the substrate specificity of an enzyme.[210] Tyrosine phosphate-induced protein interactions involving SH2 domains[213] have developed a common mechanism for stabilizing protein complexes that perform complex biological functions, particularly in intracellular signal transduction.[214−217] Protein phosphorylation has also been shown to be involved in targeting the protein IκB for ubiquitin-mediated destruction[218] to prolong the half-life of the yeast protein SST-2[219] and to control the DNA binding and transactivating activities of transcriptional activators.[220] This list of biological functions of protein phosphorylation is by no means complete, and additional functions controlled by protein phosphorylation are likely to yet be discovered. However, this overview illustrates some of the main challenges of proteome-wide phosphorylation studies, such as the large number of proteins phosphorylated, the diverse physiological conditions that induce the phosphorylation of specific groups of proteins in a cell, and the determination of the functional significance of an observed phosphorylation event.

b. Process of Protein Phosphorylation Analysis. The principal aims of essentially any protein phosphorylation study are 3-fold (Figure 6). The first is to determine the amino acid residues that are phosphorylated in vivo in a protein present in a cell in a given biological state. The second is to identify the kinase/phosphatase(s) responsible for catalyzing the specific reaction. The third is to understand the functional significance of the observed phosphorylation events for the biology of the cell. Among these aims, only the first can be directly addressed by MS and will be covered further. It is apparent, however, that the proteomics technologies discussed in the first part of this paper and other systematic or global research methods are playing an increasingly significant role in the comprehensive analysis of the

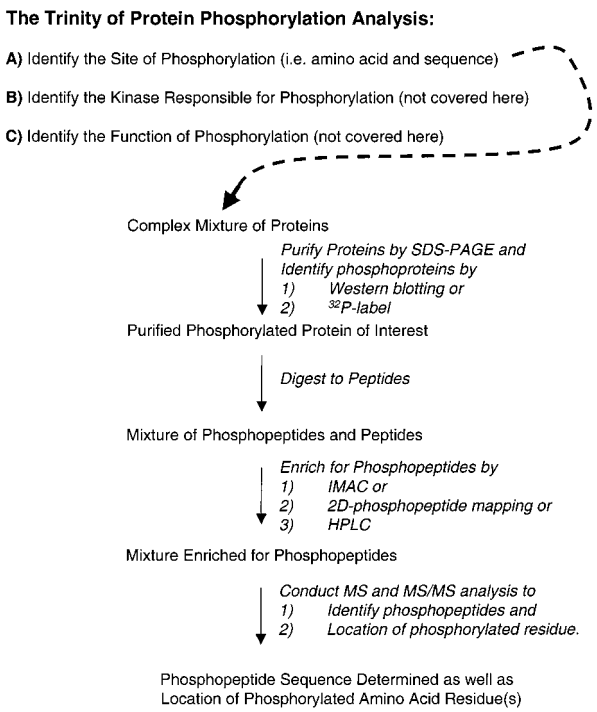


Figure 6. Phosphoprotein analysis scheme. Examples of some of the many possible routes for identifying phosphorylated amino acids in a protein and the local amino acid sequence around the phosphorylated residue.

biology of protein phosphorylation. For example, the systematic, quantitative comparison of the abundance of a large number (ideally all) of proteins expressed in a wild-type cell and in a cell expressing a regulatory protein with the site(s) of phosphorylation mutated might shed light on the processes that are controlled by the regulatory protein in question. Likewise, the availability of all the 123 currently recognized protein kinases of the S. *cerevisiae* proteome in the form of GST-fusion proteins overexpressed one at a time in genetically engineered yeast strains should make the identification of the kinase(s) that can potentially phosphorylate a particular site in a target protein relatively straightforward.[221]

MS would most efficiently identify the precise residues phosphorylated in a proteome if phosphorylated peptides could be selectively isolated from a digest of the proteins contained in a sample, separated, and analyzed by tandem mass spectrometry. Although some progress toward a method[222,223] with these properties has been achieved, no proteome-wide phosphorylation studies have been reported to date. Essentially all the experiments reported to date that involve the identification of phosphorylation sites first attempt to purify the phosphorylated protein to homogeneity before the protein is enzymatically fragmented and the phosphopeptides are isolated and mass spectrometrically analyzed. Since proteins are frequently phosphorylated to low stoichiometry (i.e., only a small fraction of a given protein may be phosphorylated) and at multiple sites (giving rise to differentially phosphorylated forms of the same protein), it is frequently difficult to isolate quantities of in vivo phosphorylated proteins that are sufficient for analysis by even the most sensitive MS methods.

Many protein phosphorylation studies are therefore performed with proteins modified by in vitro kinase reactions that can generally be scaled up to produce larger amounts of the phosphorylated protein. However, before sites of phosphorylation determined by in vitro studies can be accepted as biologically significant, the occurrence of the same phosphorylation sites in vivo needs to be established. This is frequently accomplished by comparing two-dimensional phosphopeptide (2DPP) maps of the same protein phosphorylated in vivo and in vitro.[224] Described in greater detail below, the technique essentially allows confirmation that the easily produced in vitro phosphoprotein can be used as a surrogate for the in vivo protein. Comigration of phosphopeptides generated by in vivo and in vitro phosphorylation, respectively, of the same protein is then taken as an indication that the same site is phosphorylated in vivo and in vitro[215,224] and that the in vitro system can be used as a source for the generation of peptides for mass spectrometric analysis.

Whether the site of phosphorylation is determined directly from a protein phosphorylated in vivo or a protein phosphorylated by an in vitro kinase reaction, the actual determination of the phosphorylated residue(s) generally consists of the following steps: (i) detection and purification of the phosphoprotein, (ii) enzymatic or chemical cleavage of the phosphoprotein into peptides, (iii) isolation of the phosphopeptides from nonphosphorylated peptides or at least phosphopeptide enrichment, and (iv) characterization of the phosphopeptides by MS. These steps are described in greater detail below.

2. Detection and Purification of Phosphoproteins

If a total cell lysate or any other complex protein mixture is analyzed, it is not a priori apparent which proteins, if any, are phosphorylated. A protein may first be suspected of being phosphorylated due to the presence of a protein band that migrates slightly slower in SDS−PAGE than the protein being studied. However, observation of two closely migrating bands in a one-dimensional gel or the observation of an array of spots of similar molecular mass but different isoelectric points in a two-dimensional polyacrylamide gel is insufficient to identify a particular protein as a phosphoprotein. It simply provides a suggestion that phosphorylation is a possible cause for the change in electrophoretic mobility. Stronger evidence that the slower migrating band may contain a phosphorylated form of the protein may be obtained if before electrophoresis the protein mixture was subjected to in vivo or in vitro labeling with ^{32}P followed by detection of the bands containing ^{32}P labeled proteins by autoradiography or storage phosphorimaging. Metabolic or in vivo radiolabeling is accomplished by incubating cells or tissue with $^{32}PO_4$ for a period long enough to equilibrate the cellular ATP pool with ^{32}P so that protein kinases can phosphorylate substrates. Protein phosphorylation in vitro is generally performed by kinase reactions using [γ-^{32}P]ATP as the source of the radiolabel and crude fractionated cell lysates or purified kinases as the source of the kinase activity. If labeling of the proteins with radioisotopes is not an option, the use of antibodies to detect phosphorylated proteins after electrophoretic separation (i.e., Western blotting) is a possible alternative method. This nonradioactive approach has been most successful for detection of tyrosine phosphorylated proteins.[225] A panel of tyrosine phosphate specific and very sensitive antibodies (4g10, py20) have been developed that have been invaluable tools for studying tyrosine phosphorylation.[226] The development of antibodies with specificity for phosphoserine or phosphothreonine[227,228] has been less successful, and reports using these reagents for the detection of proteins phosphorylated at these sites have been less frequent.[229] If sufficient amounts of the phosphoprotein are present in the gel for mass spectrometric detection, the bands can be excised, digested with trypsin, and subjected to data-dependent tandem MS as described above. We have recently described simple guidelines to estimate the amount of phosphoprotein present in a sample and to assess the chances for success of mass spectrometric analysis.[230] Most commonly, the phosphorylated peptides represent minor components in the peptide sample and need to be enriched prior to MS analysis in order to raise their signal above the level of the general background of low-intensity ions. This is especially true when data-dependent MS protocols are used that select ions for CID based on their relative signal intensities in the MS scan or based on a predetermined signal threshold.

3. Phosphopeptide Separation Methods

Among the common separation techniques, two-dimensional (electrophoresis/TLC) phosphopeptide mapping on cellulose plates (2DPP),[230] RP-HPLC,[231] 1D and 2D high-resolution gel electrophoresis,[232] immobilized metal affinity chromatography (IMAC),[233] and capillary electrophoresis[224] have been successfully used for the separation of phosphopeptides. Peptide separation techniques help to concentrate phosphopeptides and therefore increase the signal-to-noise ratio. Reproducible patterns of separated, radiolabeled phosphopeptides can also be used to quantitatively determine changes in the phosphorylation state of a protein as a function of time or cellular state, provided that a quantitative method is available for their detection. Peptide separation methods also effectively remove nonpeptidic contaminants, thus facilitating the detection and analysis of low-abundance phosphopeptides. Each one of the five methods (2DPP mapping, RP-HPLC, high-resolution gel electrophoresis, IMAC, and capillary electrophoresis) is compatible with further mass spectrometric analysis of the separated peptides. Finally, each of the separation methods can be used to calculate an absolute quantity of purified phosphopeptides if they are radiolabeled by in vitro kinase reactions to a known specific activity.[230] Knowing the amount of purified phosphopeptide is critical for the choice of a suitable MS strategy and for an assessment of the chances for success of the experiment.

a. Two-Dimensional Phosphopeptide Mapping. In 2DPP mapping, peptides are separated in a first dimension by electrophoresis on a thin-layer

cellulose plate and in the second dimension by thin-layer chromatography (TLC) on the same plate.[234] Separated, ^{32}P-radiolabeled phosphopeptides are then detected by autoradiography or storage phosphor imaging. The method provides important qualitative and quantitative data about the phosphorylation state of the protein not available from any other method. These include the following: (i) the maximum number of phosphorylation sites as maps generally produce more spots than there are phosphorylation sites because of differential processing by proteases. (ii) The relative stoichiometry of phosphorylation among all phosphopeptides is provided by autoradiographic intensity. (iii) The relative hydropathy of the separated phosphopeptides is apparent from the tangential separations of electrophoresis and TLC. A significant advantage of 2DPP mapping is that it produces purified phosphopeptides that can be analyzed, after extraction from the plate, directly by MS methods.[235] If 2DPP mapping is intended as a preparative method for MS/MS analysis, then the amount of protease added should be kept as low as is practical. Use of excess protease will be obvious when analyzing a "purified" phosphopeptide by MS because autocatalytic products from the protease will dominate the spectra and may prevent analysis of the phosphopeptide. Furthermore, the 2DPP method is very sensitive and can be even more sensitive than MS methods because detection is by integration of radioactive decay over potentially very long time periods. Finally, the high degree of pattern reproducibility achieved by the method makes 2DPP the method of choice for projects in which the state of phosphorylation of a protein under different conditions needs to be analyzed using, for example, time courses and different induced states of activation.

b. High-Resolution Gel Electrophoresis. Preparative methods using 1D or 2D electrophoresis to purify phosphopeptides on polyacrylamide gels, respectively, were recently published.[232] In the 2D method, nondenaturing gel isoelectric focusing was combined with alkaline 40% polyacrylamide gel electrophoresis for phosphopeptide separation and comparative pattern analysis as is done with 2DPP mapping. Phosphopeptides were detected by autoradiography or storage phosphorimaging of ^{32}P-labeled samples. Edman sequencing rather than MS was used to identify the proteins and to determine the sites of phosphorylation, but the method is presumably adaptable to MS-based methods. The method promises the same results as 2DPP mapping except that the recovered samples might be less contaminated with the non-peptide components that compete with peptide analytes for ionization in the mass spectrometer and that are carried along with the peptides after extraction from the cellulose matrix used for 2DPP mapping. Unlike with 2DPP mapping, there is the potential for loss of specific phosphopeptides if electrophoresis is not closely monitored. Regardless, the method is appealing because it uses relatively common equipment whereas 2DPP mapping requires purchase of specialized equipment for conducting the first dimension electrophoretic separation.

c. Ion Metal Affinity Chromatography. One commonly overlooked difficulty with phosphopeptide analysis is the low stoichiometry of phosphorylation. In such cases phosphopeptide(s) are present in the sample in very small amounts as compared to the nonphosphorylated peptide with the same sequence and the other peptides derived from the digested protein. It is thus difficult to identify phosphopeptides by MS techniques even though their presence in the sample is confirmed by ^{32}P label that was detected in a 2DPP spot, a whole protein digest, or a HPLC fraction. As mentioned before, data-dependent tandem MS methods often fail to identify minor species in a sample because priority for selection for CID goes to the most intense ion detected in the MS scan. To alleviate this problem, selective enrichment of phosphopeptides by IMAC can be employed.[224,233] The technique involves chelation of metals such as Fe^{3+} or Ga^{3+} onto a chromatographic support consisting of iminodiacetic acid or nitrilotriacetic acid.[236−238] Phosphopeptides, being acidic by virtue of the phosphate group, bind with some selectively over nonphosphopeptides. Fractions enriched for phosphopeptides are then eluted by phosphate or increased pH. While the method is somewhat selective for phosphopeptides, other peptides, particularly those containing strings of acidic amino acids or histidine, are also enriched. The method has been applied on-line to MS in an integrated peptide enrichment/separation system consisting of a tandem IMAC/RP column configuration.[215,224,235]

d. Reversed-Phase HPLC. Reversed-phase HPLC fractionation of phosphopeptides is reproducible, simple, and does not require specialized equipment.[225,231,239] In RP-HPLC, ^{32}P-labeled phosphopeptides are separated on the basis of their hydrophobicity and fractions collected for Cerenkov counting (i.e., without scintillation fluid so that the samples may be further analyzed by MS). A graph of Cerenkov counts versus elution time reveals the number of radioactive fractions that can then be analyzed by the MS methods described below.[231,239] A disadvantage of RP-HPLC over 2DPP mapping is that very hydrophilic phosphopeptides may not stick to the column and thus will elute in the column flow-through. Conversely, very hydrophobic peptides may not elute until the end of a gradient and will be obscured by the polymeric contaminants that often elute at high acetonitrile concentrations or they may not elute at all. It is therefore possible that some of the phosphopeptides in a sample will go undetected. Generally, the resolution of RP-HPLC is also inferior to the resolution achieved by the two-dimensional peptide mapping technique. An additional note of caution is that phosphopeptides will stick to metal surfaces. Significant sample losses can occur if standard metal injectors are used. Even considering these disadvantages, RP-HPLC for phosphopeptide analysis is popular because of the ease with which RP-HPLC systems are connected on-line to ESI mass spectrometers. The use of a mass spectrometer connected on-line to the HPLC system makes it possible to detect and characterize phosphopeptides in sample mixtures, even if the analyte is not radiolabeled. This

is achieved by implementing one of several possible phosphate-specific diagnostic ion scans, which include precursor ion scans,[240] neutral loss scans,[241] and in-source dissociation.[242]

e. Capillary Electrophoresis. Recently, two methods using capillary electrophoresis (CE) for analysis of phosphopeptides via ESI were published.[141,224] The method by Figeys et al. incorporated a solid-phase extraction (SPE) capillary zone electrophoresis (CZE) device for peptide concentration/separation on-line with ESI-MS and an algorithm written in Instrument Control Language (ICL) that modulated the electrophoretic conditions in a data-dependent manner to optimize available time for the generation of high-quality CID spectra of peptides in complex samples. The data-dependent modulation of the electric field significantly expanded the analytical window for each peptide analyzed and enhanced the sensitivity by reducing the CE voltage and thus the flow into the ESI source. The technique was applied to the analysis of in vivo phosphorylation sites of endothelial nitric oxide synthase (eNOS) demonstrating the power of the method for the MS/MS analysis of minor peptide species in complex samples such as phosphopeptides generated by the proteolytic digestion of a large protein, eNOS, phosphorylated at low stoichiometry. The second application of CE for analysis of phosphopeptides took a slightly different approach technically but also was concerned with analysis of minor phosphopeptide components produced by proteolysis of large proteins.[243] The method used Fe(III) immobilized metal-ion affinity chromatography (IMAC)-CE-electrospray ionization MS to analyze subpicomole analysis of phosphopeptides. The IMAC resin was packed directly at the head of the CE column, and after the phosphopeptides were bound to the resin and washed, they were eluted at high pH and separated by CE. Advantages of this approach include (i) selective retention and preconcentration of phosphopeptides; (ii) a prewash of the sample to remove salts and buffers that are not suited for CE separation or ESI operation; (iii) ease of construction; and (iv) adaptation to commercial CE instruments without any modifications.

4. Phosphopeptide Sequence Determination

There are number of different mass spectrometric methods for determining which amino acid residue(s) in a peptide are phosphorylated, and they fall into two general themes. The first relies on the chemical lability of the phosphoester bonds in phosphoserine, -threonine, and -tyrosine. These phosphoester bonds can easily be induced to fragment in a collision cell or the ion source of an ESI instrument or during PSD in a MALDI-MS, resulting in loss of phosphate from the peptide. Phosphopeptides that lose phosphate due to any of these processes can then be identified by any of several phosphate-specific diagnostic ion scans. The second theme relies on the detection of the mass added to a peptide by the phosphate group. Typically, in protein phosphorylation studies, the amino acid sequence of the protein investigated is known. Therefore, phosphopeptides derived from the protein can, in principle, be detected by a net mass differential of 80 u that occurs when phosphate is added to serine, threonine, or tyrosine. Thus, a peptide mass map of the proteolytically fragmented phosphoprotein can potentially identify the phosphorylated peptide by comparison to the known protein sequence. Neither method however identifies the phosphorylated amino acid residue(s) within the peptide directly. In cases where the peptide sequence contains only a single possible phosphorylation site, the phosphorylated residue is effectively located by default. If this is not the case, then tandem MS is necessary to locate the phosphorylated amino acid residue. In general, methods that produce some sort of phosphate-specific ion (i.e., a diagnostic ion) are useful or even essential for the detection of phosphopeptides in mixtures in cases in which incorporation of ^{32}P is not possible or in which the radiolabel has decayed past the point of detection.[244] However, such scans can also be used on radiolabeled phosphopeptides because the contribution to the mass of the phosphopeptide from the radioactive isotope of phosphate is so small that it can be ignored. If researchers are concerned about contaminating a mass spectrometer with ^{32}P samples, they can easily avoid contamination by waiting a sufficient number of half-lives (2 weeks for ^{32}P) before conducting mass spectrometric experiments. Several types of MS-based approaches to phosphopeptide analysis have been developed and applied. The most commonly used ones are described in the following.

a. In-Source CID. If phosphopeptide ions are fragmented in negative ion mode, $H_2PO_4^-$ (97 u), PO_3^- (79 u), and PO_2^- (63 u) are detected as phosphate-specific diagnostic ions. Under low-energy CID conditions, phosphotyrosine will be observed to generate the last two of these three ions, PO_3^- (79 u) and PO_2^- (63 u) but not $H_2PO_4^-$ (97 u). These phosphate-specific diagnostic ions can be selectively monitored to identify phosphopeptides.[245,246] When in-source CID is combined on-line with HPLC, a chromatographic trace is established that identifies the elution time of a phosphopeptide. Carr and coworkers developed a negative ion LC-MS protocol that monitors the phosphopeptide-specific reporter ions and determines the phosphopeptide molecular weight in the same scan.[245] This was accomplished by use of a high orifice potential across the two skimmers prior to Q1 in a triple quadrupole instrument while the low *m*/*z* range is scanned for the diagnostic ions. The orifice potential is then lowered to a voltage that does not induce fragmentation and the high *m*/*z* range is scanned. A similar experiment can be done on instruments where a heated capillary replaces the first skimmer.[246,247] This method, an extension of that of Hunter and Games,[246] uses an alternating scan approach where selected ion monitoring of appropriate diagnostic ions at a high octapole offset voltage is followed by two full scans. The first full scan is conducted at the same high offset voltage as the SIM experiment providing signals for the deprotonated phosphopeptide molecular ion and the phosphopeptide molecular ion minus phosphate. Finally a second full scan is done at a normal octapole offset voltage to provide a reference to the full scan at high octapole offset. This series of three MS scans

is repeated continuously throughout the LC separation. Such an experiment provides the same information as the methods of Carr and co-workers,[245] but because selected ion monitoring (SIM) is used rather than scanning to detect diagnostic ions, one cycle of scans is faster and potentially more sensitive. The first full scan at high octapole offset is compared to the full scan at low octapole offset to provide a clue as to which peptide ion is phosphorylated. This is useful to distinguish phosphopeptides from nonphosphorylated peptides in case they coelute from the microcapillary columns. Such techniques are generally capable of identifying low femtomole amounts of standard applied to a capillary column. Frequently, the detection sensitivity with "real" in vitro or in vivo phosphorylated samples is lower and can drop into the picomole range.

Of course it would be advantageous to also determine the amino acid sequence of the detected phosphopeptide and the phosphorylated residue(s) in the same negative ion LC-MS experiment. Unfortunately, this has been difficult to achieve because negative ion CID spectra generally produce insufficient fragment ions for sequence elucidation. Attempts to switch between negative ion mode for phosphopeptide detection and positive ion mode for peptide product ion scanning in the same LC-MS experiment have been technically difficult in scanning mass spectrometers (such as quadrupole instruments) due to the time required to switch between positive and negative ion mode in real time. It appears that such experiments might be easier to carry out in nonscanning mass spectrometers.

b. Neutral Loss Scanning. Neutral loss scanning carried out in positive ion mode with ESI in a triple quadrupole MS for phosphopeptide detection and analysis was first described by Covey et al.[241] and further developed by Huddelston et al.[245] Instead of using Q1 to select specific ions for fragmentation in Q2, Q1 is scanned coordinately with Q3 but with an offset in *m*/*z* value. The offset in the *m*/*z* values measured by the two quadrupoles at any given time is constant and corresponds to the loss of phosphate by CID in Q2. This loss reflects the facile loss of phosphate from phosphoserine and phosphothreonine due to a process known as β-elimination. Only phosphothreonine and phosphoserine (not shown) may undergo neutral loss of H_3PO_4 (98) by β-elimination as shown in Figure 7 but not phosphotyrosine.[248] This is because the α-carbon proton that is abstracted by a lone pair of electrons from the phosphate moiety is too far removed for facile loss (Figure 7B); i.e., in tyrosine phosphate the phosphate group is no longer in the β position relative to the α-carbon proton. Therefore, neutral loss scanning for a loss of phosphate from a $[M + 2H]^{2+}$ phosphopeptide ion requires an offset value is 49 *m*/*z*. The method has not been as popular as the aforementioned in-source CID methods because of the propensity of false-positive signals and the need to know the charge state of the ion losing phosphate. It is an advantage of the method that it is carried out in positive ion mode and can therefore be used with data-dependent scanning to acquire CID spectra in the same experiment using the detection of a neutral loss of phosphate as a trigger to initiate CID.

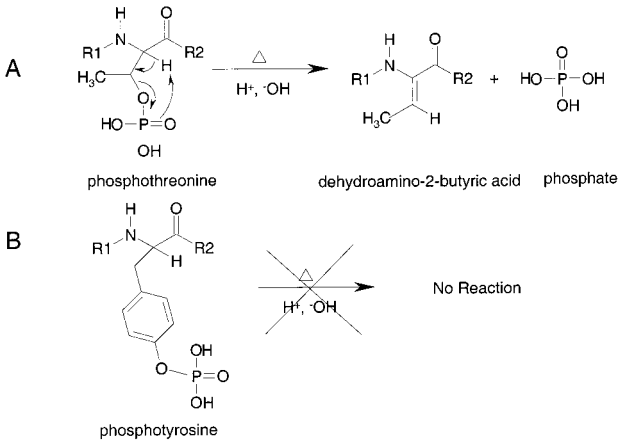


Figure 7. Mechanism for β-elimination of phosphate. Illustration of chemical mechanism for (A) β-elimination of phosphate from phosphothreonine to produce dehydroamino-2-butyric acid (scheme is also valid for phosphoserine where dehydroalanine forms) and phosphate and (B) absence of the same mechanistic pathway for β-elimination of phosphate from phosphotyrosine.

c. Precursor Ion Scanning. In this method, negative ion ESI is carried out with continuous scanning of Q1 and Q3 maintained at a fixed *m*/*z* ratio. Ions are fragmented in Q2, and Q3 passes only one ion; in the case of phosphopeptide analysis this is usually 79 *m*/*z* (i.e., loss of PO_3^-). Consequently, the resultant mass spectrum shows only ions that lost 79 *m*/*z*.[140,249,250] This greatly simplifies mixture analysis and is best done during direct infusion with a nanospray source. Again as described for the phosphate diagnostic ion scans, there is a problem associated with sequencing in positive ion mode immediately after detecting the loss of phosphate in negative ion mode.

d. Product Ion Scanning. Often the information obtained by the specific scanning methods described above is not sufficient for identification of the phosphorylated residue in a phosphopeptide. In fact, the above methods of in-source CID, neutral loss, and precursor ion scanning are designed to distinguish phosphopeptides from nonphosphopeptides and to potentially indicate the phosphopeptide mass rather than to provide sequence information. Consequently, these methods can only successfully identify a phosphorylated residue if the peptide sequence is known and contains only one copy of one of the hydroxyl amino. If there is more than one possible amino acid residue present in the peptide that can be phosphorylated, then it is necessary to acquire tandem mass spectra for either manual or algorithm-based sequence interpretation.[231,239]

As a general trend for low-energy CID of phosphopeptides, it has been observed that phosphate tends to be lost from phosphoserine more readily than phosphothreonine and from phosphothreonine more readily than from phosphotyrosine. Phosphate is generally eliminated from shorter phosphopeptides more readily than from longer phosphopeptides because roughly the same amount of energy for collision is dispersed across fewer bonds. Interestingly, it is

rare to observe the immonium ions for phosphoamino acids that form as a result of dehydroalanine and dehydroamino-2-butyric acid breaking down after loss of phosphate. However, using an ion trap mass spectrometer and monitoring the CID of a phosphopeptide ion, dehydroamno-2-butyric acid (Figure 7A) was observed in place of threonine in the peptide fragment ion.[250]

e. Post-Source Decay. Meta-stable decay of phosphopeptides has been observed during PSD-MALDI-TOF and provides a method to sequence peptides in a single-stage instrument. While not popular for reasons cited above, the method has been successfully applied to the analysis of phosphopeptides.[236,251]

f. Enzymatic Dephosphorylation. Phosphatases can be used to identify phosphopeptides in a mixture of predominantly nonphosphopeptides. Typically, as a first step, the peptide masses resulting from proteolytic digestion of the phosphoprotein are acquired in a MALDI-TOF instrument. Second, the same sample is treated with phosphatase to remove phosphate selectively from the phosphopeptide(s) and the masses acquired again. Any peptide mass that decreases by 80 u as a result of the phosphatase treatment will be designated a potential phosphopeptide. An advantage to conducting such an experiment by MALDI-MS is that peptide ions produced tend to be singly charged rather than multiply protonated and that the phosphatase reaction can be carried out directly on the MALDI probe.[252,253] A similar approach has also been developed for ESI-MS.[254] In this method, a enzyme microreactor consisting of an immobilized tyrosine phosphatase was used to dephosphorylate peptides on-line prior to analysis by CE-MS/MS. As for the MALDI-based method, phosphopeptides in the peptide mixture were characterized by a mass difference of 80 u when the MS data obtained with and without enzyme reactor were compared. The method has the additional advantage that the peptides were also undergoing a mobility shift in CE upon dephosphorylation, further confirming their identity as phosphopeptides and that the phosphorylated or dephosphorylated species of the phosphopeptide could be directly subjected to CID for further characterization and location of the phosphorylated residue, respectively.

IV. Conclusions

Traditionally, advances in mass spectrometric methods for the analysis of proteins and peptides were driven primarily by the need to identify and analyze purified proteins faster, more sensitively, and more reliably. The advent of complete genome sequences has accelerated incremental improvements in mass spectrometric methods for protein identification and analysis and also catalyzed a new research method.

Incremental improvements have been accelerated because the genome sequence databases contain the sequence information for every protein potentially expressed by that genome. Consequently, proteins isolated from species with complete sequence databases are no longer identified by de novo sequencing but rather by correlating idiotypic information extracted from the intact polypeptide or a peptide fragment thereof with the sequence database. Currently, MS is the method of choice for the generation of data for sequence database searching and therefore a cornerstone of analytical protein chemistry.

The genomics revolution has also catalyzed a new research method we have termed discovery science.[4] Discovery science enumerates the elements of a biological system irrespective of any hypotheses of how the system functions. Discovery science complements the traditional hypothesis driven method to biological research, and proteomics is an essential component of discovery science. The initial efforts of proteomics have been focused on the identification of the proteins expressed by a cell or tissue a process that can be described as descriptive proteomics. More recently, the focus has shifted to the development of methods capable of measuring, on a proteome-wide scale, properties of proteins that reflect the function and dynamics of proteins. These include the quantity, the state of modification, the specific activity, and the association of a protein with other macromolecules. Many of these methods depend on MS and are currently being rapidly further developed. The potential for method refinement, for developing methods to uncover new types of information, and for the power of the current methods to dissect biological systems at a molecular level make MS and proteomics among the most exciting, dynamic, and important research themes at the present time.

Acknowledgments. D.R.G. thanks Dr. J. D. Watts for critical opinions and edification in protein phosphorylation analysis over the last 3 yr. More information about the author's research can be found at http://www.systemsbiology.org. For issues concerning MS, the American Society for Mass Spectrometry has an excellent website at http://www.asms.org.

V. References

(1) Rowen, L.; Mahairas, G.; Hood, L. *Science* **1997**, *278*, 605.
(2) Fraser, C. M.; Fleischmann, R. D. *Electrophoresis* **1997**, *18*, 1207.
(3) Harry, J. L.; Wilkins, M. R.; Herbert, B. R.; Packer, N. H.; Gooley, A. A.; Williams, K. L. *Electrophoresis* **2000**, *21*, 1071.
(4) Aebersold, R.; Hood, L. E.; Watts, J. D. *Nat. Biotechnol.* **2000**, *18*, 359.
(5) Pandey, A.; Mann, M. *Nature* **2000**, *405*, 837.
(6) Goffeau, A.; Barrell, B. G.; Bussey, H.; Davis, R. W.; Dujon, B.; Feldmann, H.; Galibert, F.; Hoheisel, J. D.; Jacq, C.; Johnston, M.; Louis, E. J.; Mewes, H. W.; Murakami, Y.; Philippsen, P.; Tettelin, H.; Oliver, S. G. *Science* **1996**, *274*, 563.
(7) Consortium. *C. Elegans, Sci.* **1998**, *282*, 2012.
(8) Adams, M. D.; Celniker, S. E.; Holt, R. A.; Evans, C. A.; Gocayne, J. D.; Amanatides, P. G.; Scherer, S. E.; Li, P. W.; Hoskins, R. A.; Galle, R. F.; George, R. A.; Lewis, S. E.; Richards, S.; Ashburner, M.; Henderson, S. N.; Sutton, G. G.; Wortman, J. R.; Yandell, M. D.; Zhang, Q.; Chen, L. X.; Brandon, R. C.; Rogers, Y. H.; Blazej, R. G.; Champe, M.; Pfeiffer, B. D.; Wan, K. H.; Doyle, C.; Baxter, E. G.; Helt, G.; Nelson, C. R.; Gabor-Miklos, G. L.; Abril, J. F.; Agbayani, A.; An, H. J.; Andrews-Pfannkoch, C.; Baldwin, D.; Ballew, R. M.; Basu, A.; Baxendale, J.; Bayraktaroglu, L.; Beasley, E. M.; Beeson, K. Y.; Benos, P. V.; Berman, B. P.; Bhandari, D.; Bolshakov, S.; Borkova, D.; Botchan, M. R.; Bouck, J.; Brokstein, P.; Brottier, P.; Burtis, K. C.; Busam, D. A.; Butler, H.; Cadieu, E.; Center, A.; Chandra, I.; Cherry, J. M.; Cawley, S.; Dahlke, C.; Davenport, L. B.; Davies, P.; de Pablos, B.; Delcher, A.; Deng, Z.; Mays, A. D.; Dew, I.; Dietz, S. M.; Dodson, K.; Doup, L. E.; Downes, M.; Dugan-Rocha, S.; Dunkov, B. C.; Dunn, P.; Durbin, K. J.; Evangelista, C. C.; Ferraz, C.; Ferriera, S.; Fleischmann, W.; Fosler, C.; Gabrielian, A. E.; Garg, N. S.; Gelbart, W. M.; Glasser, K.; Glodek, A.; Gong, F.; Gorrell, J. H.; Gu, Z.; Guan, P.; Harris, M.; Harris, N. L.; Harvey, D.; Heiman, T. J.; Hernandez, J. R.; Houck, J.; Hostin, D.; Houston, K. A.; Howland, T. J.; Wei, M.

H.; Ibegwam, C.; Jalali, M.; Kalush, F.; Karpen, G. H.; Ke, Z.; Kennison, J. A.; Ketchum, K. A.; Kimmel, B. E.; Kodira, C. D.; Kraft, C.; Kravitz, S.; Kulp, D.; Lai, Z.; Lasko, P.; Lei, Y.; Levitsky, A. A.; Li, J.; Li, Z.; Liang, Y.; Lin, X.; Liu, X.; Mattei, B.; McIntosh, T. C.; McLeod, M. P.; McPherson, D.; Merkulov, G.; Milshina, N. V.; Mobarry, C.; Morris, J.; Moshrefi, A.; Mount, S. M.; Moy, M.; Murphy, B.; Murphy, L.; Muzny, D. M.; Nelson, D. L.; Nelson, D. R.; Nelson, K. A.; Nixon, K.; Nusskern, D. R.; Pacleb, J. M.; Palazzolo, M.; Pittman, G. S.; Pan, S.; Pollard, J.; Puri, V.; Reese, M. G.; Reinert, K.; Remington, K.; Saunders: R. D.; Scheeler, F.; Shen, H.; Shue, B. C.; Siden-Kiamos, I.; Simpson, M.; Skupski, M. P.; Smith, T.; Spier, E.; Spradling, A. C.; Stapleton, M.; Strong, R.; Sun, E.; Svirskas, R.; Tector, C.; Turner, R.; Venter, E.; Wang, A. H.; Wang, X.; Wang, Z. Y.; Wassarman, D. A.; Weinstock, G. M.; Weissenbach, J.; Williams, S. M.; Woodage, T.; Worley, K. C.; Wu, D.; Yang, S.; Yao, Q. A.; Ye, J.; Yeh, R. F.; Zaveri, J. S.; Zhan, M.; Zhang, G.; Zhao, Q.; Zheng, L.; Zheng, X. H.; Zhong, F. N.; Zhong, W.; Zhou, X.; Zhu, S.; Zhu, X.; Smith, H. O.; Gibbs, R. A.; Myers, E. W.; Rubin, G. M.; Venter, J. C. *Science* **2000**, *287*, 2185.
(9) Butler, D.; Pockley, P. *Nature* **2000**, *404*, 534.
(10) Pennisi, E. *Science* **2000**, *288*, 239.
(11) Pennisi, E. *Science* **2000**, *289*, 2304.
(12) National Cancer Institute. dbEST, 2000, www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html.
(13) Hofmann, K.; Bucher, P.; Falquet, L.; Bairoch, A. *Nucleic Acids Res.* **1999**, *27*, 215.
(14) Henikoff, S.; Henikoff, J. G.; Pietrokovski, S. *Bioinformatics* **1999**, *15*, 471.
(15) Skolnick, J.; Fetrow, J. S.; Kolinski, A. *Nat. Biotechnol.* **2000**, *18*, 283.
(16) Marcotte, E. M.; Pellegrini, M.; Ng, H. L.; Rice, D. W.; Yeates, T. O.; Eisenberg, D. *Science* **1999**, *285*, 751.
(17) Enright, A. J.; Iliopoulos, I.; Kyrpides, N. C.; Ouzounis, C. A. *Nature* **1999**, *402*, 86.
(18) Wilkins, M. R.; Sanchez, J. C.; Gooley, A. A.; Appel, R. D.; Humphery-Smith, I.; Hochstrasser, D. F.; Williams, K. L. *Biotechnol. Genet. Eng. Rev.* **1996**, *13*, 19.
(19) Wasinger, V. C.; Cordwell, S. J.; Cerpa-Potjak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R., Duncan, M. W.; Harris, R. Williams, K. L.; Humphrey-Smith, I. *Electrophoresis* **1995**, *16*, 1090.
(20) Hochstrasser, D. F. *Clin. Chem. Lab. Med.***1998**, *36*, 825.
(21) Loo, R. R. O.; Stevenson, T. I.; Mitchell, C.; Loo, J. A.; Andrews, P. C. *Anal. Chem.* **1996**, *8*, 1910.
(22) Haynes, P. A.; Gygi, S. P.; Figeys, D.; Aebersold, R. *Electrophoresis* **1998**, *19*, 1862.
(23) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.
(24) Cole, R. B. *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation and Applications*; Wiley: New York, 1997.
(25) Karas, M.; Hillenkamp, F. *Anal. Chem.* **1995**, *60*, 2299.
(26) Barber, M.; Bordoli, R. S.; Sedgwick, R. D.; Tyler, A. N. J. *Chem. Soc. Commun.* **1981**, 325.
(27) Jardine, I. *Methods Enzymol.* **1990**, *193*, 441.
(28) Chaurand, P.; Luetzenkirchen, F.; Spengler, B. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 91.
(29) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Anal. Chem.* **1993**, *65*, 425−38.
(30) Chait, B. T.; Kent, S. B. *Science* **1992**, *257*, 1885.
(31) Loo, J. A. *Bioconjugate Chem.* **1995**, *6*, 644.
(32) Goodlett, D. R.; Ogorzalek-Loo, R. R.; Loo, J. A.; Wahl, J. H.; Udseth, H. R.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 614.
(33) Ganem, B.; Li, Y.-T.; Henion, J. D. *J. Am. Chem. Soc.* **1991**, *113*, 7818.
(34) Loo, J. A. *Mass Spectrom. Rev.* **1997**, *16*, 1.
(35) Wood, T. D.; Chorush, R. A.; Wampler, F. M., III; Little, D. P.; O'Connor, P. B.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2451.
(36) Anderegg, R. J.; Wagner, D. S.; Stevenson, C. L.; Borchardt, R. T. *J. Am. Soc. Mass Spectrom.* **1994**, *4*, 425.
(37) Akashi, S.; Naito, Y.; Takio, K. *Anal. Chem.* **1999**, *71*, 4974.
(38) Loo, J. A.; Loo, R. R.; Udseth, H. R.; Edmonds, C. G.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 101.
(39) Katta, V.; Chait, B. T. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 214.
(40) Mirza, U. A.; Cohen, S. L.; Chait, B. T. *Anal. Chem.* **1993**, *65*, 1.
(41) Konermann, L.; Collings, B. A.; Douglas, D. J. *Biochemistry* **1997**, *36*, 5554.
(42) Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5802.
(43) Cohen, S. L.; Padovan, J. C.; Chait, B. T. *Anal. Chem.* **2000**, *72*, 574.
(44) Rappsilber, J.; Siniossoglou, S.; Hurt, E. C.; Mann, M. *Anal. Chem.* **2000**, *72*, 267.
(45) Kennedy, R.; J. W. Jorgenson, J. W. *Anal. Chem.* **1991**, *63*, 1467.
(46) Hunt, D. F.; Alexander, J. E.; McCormack, A. L.; Martino, P. A.; Michel, H.; Shabanowitz, J.; Sherman, N.; Moseley, M. A.; Jorgenson, J. W.; Tomer, K. B. *Techniques in Protein Chemistry II*; Academic Press: New York, 1991; p 441.
(47) Lee, N.; Goodlett, D. R.; Marquardt, H.; Geraghty, D. E. *J. Immunol.* **1998**, *160*, 4951.
(48) Mosely, M. A.; Deterding, L. J.; Tomer, K. B.; Jorgenson, J. W. *Anal. Chem.* **1991**, *63*, 1467.
(49) Wahl, J. H.; Goodlett, D. R.; Udseth, H. R.; Smith, R. D. *Anal. Chem.* **1992**, *64*, 3194.
(50) Wahl, J. H.; Goodlett, D. R.; Udseth, H. R.; Smith, H. R. *Electrophoresis* **1993**, *14*, 448.
(51) Wahl, J. H.; Gale, D. C.; Smith, R. D. *J. Chromatogr. A* **1994**, *659*, 217.
(52) Wilm, M. S.; Mann, M. *Int. J. Mass Spectrom. Ion Processes* **1994**, *136*, 167.
(53) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1.
(54) Hunt, D. F.; Yates, J. R., III; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233.
(55) Matsudaira, P. *J. Biol. Chem.* **1987**, *262*, 10035.
(56) Lazar, I. M.; Ramsey, R. S.; Sundberg, S.; Ramsey, J. M. *Anal. Chem.* **1999**, *71*, 3627.
(57) Carr, S. A.; Annan, R. S. *Current Protocols in Molecular Biology*; John Wiley & Sons: New York, 1997; p 10.21.1.
(58) Davis, M. T.; Lee, T. D. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 1059.
(59) McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R., III. *Anal. Chem.* **1997**, *69*, 767.
(60) Figeys, D.; Aebersold, R. *Electrophoresis* **1998**, *19*, 885.
(61) Figeys, D.; Gygi, S. P.; McKinnon, G.; Aebersold, R. *Anal. Chem.* **1998**, *70*, 3728.
(62) Belov, M. E.; Gorshkov, M. V.; Udseth, H. R.; Anderson, G. A.; Tolmachev, A. V.; Prior D. C.; Harkewicz, R.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 19.
(63) Zhang, B.; Foret, F.; Karger, B. L. *Anal.Chem.* **2000**, *72*, 1015.
(64) Yost, R. A.; Enke, C. G. *Anal. Chem.* **1979**, *51*, 1251A.
(65) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Science* **1990**, *248*, 201.
(66) Schwartz, J. C.; Jardine, I. *Methods Enzymol.* **1996**, *270*, 552.
(67) Jonscher, K. R.; Yates, J. R., III. *Anal. Biochem.* **1997**, *244*, 1.
(68) Morris, H. R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R. S.; Hoyes, J.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 889.
(69) Borchers, C.; Parker, C. E.; Deterding, L. J.; Tomer, K. B. *J. Chromatogr. A* **1999**, *854*, 119.
(70) Borchers, C.; Peter, J. F.; Hall, M. C.; Kunkel, T. A.; Tomer, K. B. *Anal. Chem.* **2000**, *72*, 1163.
(71) Fitzgerald, M. C.; Chernushevich, I. I.; Standing, K. G.; Whitman, C. P.; Kent, S. B. *Proc. Natl. Acad. Sci U.S.A.* **1996**, *93*, 6851.
(72) Tanaka, K.; Kawatoh, E.; Ding, L.; Smith, A. J.; Kumashiro, S. *Proceedings of 47th American Society for Mass Spectrometry and Allied Topics*; 1999; TP086.
(73) Medzihradszky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick A. M.; Juhasz, P.; Vestal, M. L.; Burlingame, A. L. *Anal. Chem.* **2000**, *72*, 552.
(74) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1.
(75) Wood, T. D.; Guan, Z.; Borders, C. L.; Chen, L. H.; Kenton, G. L.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3362.
(76) Qin, J.; Chait, B. T. *Anal. Chem.* **1997**, *69*, 4002.
(77) Qin, J.; Fenyo, D.; Zhao, Y.; Hall, W. W.; Chao, D. M.; Wilson, C. J.; Young, R. A.; Chait, B. T. *Anal. Chem.* **1997**, *69*, 3995.
(78) Conuterman, A. E.; Valentine, S. J.; Srebalus, C. A.; Henderson, S. C.; Hoagland, C. S.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 743.
(79) Burlingame, A. L.; Boyd, R. K.; Gaskell, S. J. *Anal. Chem.* **1996**, *68*, 599R.
(80) Burlingame, A. L.; Boyd, R. K.; Gaskell, S. J. *Anal. Chem.* **1998**, *68*, 647R.
(81) Hewick, R. M.; Hunkapiller, M. W.; Hood, L. E.; Dreyer, W. J. *J. Biol. Chem.* **1981**, *256*, 7990.
(82) Aebersold, R. H.; Leavitt, J.; Saavedra, R. A.; Hood, L. E.; Kent, S. B. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6970.
(83) Larive, C. K.; Lunte, S. M.; Zhong, M.; Perkins, M. D.; Wilson, G. S.; Gokulrangan, G.; Williams, T.; Afroz, F.; Schoneich, C.; Derrick, T. S.; Middaugh, C. R.; Bogdanowich-Knipp, S. *Anal. Chem.* **1999**, *71*, 398R.
(84) Lamond, A. I.; Mann, M. *Trends Cell Biol.* **1997**, *7*, 139.
(85) Patterson, S. D.; Aebersold, R. *Electrophoresis* **1995**, *16*, 1791.
(86) Moritz, R. L.; Eddes, J.; Ji, H.; Reid, G. E.; Simpson, R. J. *Techniques in Protein Chemistry VI*; Academic Press: San Diego, 1995; p 311.
(87) Aebersold, R.; Leavitt, J. *Electrophoresis* **1990**, *11*, 517.
(88) Aebersold, R. H.; Teplow, D. B.; Hood, L. E.; Kent, S. B. H. *J. Biol. Chem.* **1986**, *261*, 4229.
(89) Tonella, L.; Walsh, B. J.; Sanchez, J. C.; Ou, K.; Wilkins, M. R.; Tyler, M.; Frutiger, S.; Gooley, A. A.; Pescaru, I.; Appel, R. D.;

Yan, J. X.; Bairoch, A.; Hoogland, C.; Morch, F. S.; Hughes, G. J.; Williams, K. L.; Hochstrasser, D. F. *Electrophoresis* **1998**, *19*, 1960.
(90) Joubert, R.; Brignon, P.; Lehmann, C.; Monribot, C.; Gendre, F.; Boucherie, H. *Yeast* **2000**, *16*, 511.
(91) Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M. *Proc. Natl. Acad. Sci U.S.A.* **1996**, *93*, 14440.
(92) Langen, H.; Takacs, B.; Evers, S.; Berndt, P.; Lahm, H. W.; Wipf, B.; Gray, C.; Fountoulakis, M. *Electrophoresis* **2000**, *21*, 411.
(93) Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L.; Hochstrasser, D. F. *Biotechnology* **1996**, *14*, 61.
(94) Wilkins, M. R.; Gasteiger, E.; Wheeler, C. H.; Lindskog, I.; Sanchez, J. C.; Bairoch, A.; Appel, R. D.; Dunn, M. J.; Hochstrasser, D. F. *Electrophoresis* **1998**, *19*, 3199.
(95) Wilkins, M. R.; Yan, J. X.; Gooley, A. A. *Methods Mol. Biol.* **1999**, *112*, 445.
(96) Gobom, J.; Nordhoff, E.; Mirgorodskaya, E.; Ekman, R.; Roepstorff, P. *Anal. Chem.* **1999**, *71*, 919.
(97) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C. Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011.
(98) James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 58.
(99) Yates, J. R., III; Speicher, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397.
(100) Pappin, D. J. *Methods Mol. Biol.* **1997**, *64*, 165.
(101) Jensen, O. N.; Larsen, M. R.; Roepstorff, P. *Proteins Suppl.* **1998**, *2*, 74.
(102) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741.
(103) Eriksson, J. Chait, B. T.; Fenyo, D. *Anal. Chem.* **2000**, *72*, 999.
(104) Cordwell, S. J.; Wilkins, M. R.; Cerpa-Poljak, A.; Gooley, A. A.; Duncan, M.; Williams, K. L.; Humphery-Smith, I. *Electrophoresis* **1995**, *16*, 438.
(105) Fenyo, D.; Qin, J.; Chait, B. T. *Electrophoresis* **1998**, *19*, 998.
(106) Jensen, O. N.; Podtelejnikov, A.; Mann, M. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1371.
(107) Clauser, K. R.; Baker, P.; Burlingame, A. L. *Anal. Chem.* **1999**, *71*, 2871.
(108) Takach, E. J.; Hines, W. M.; Patterson, D. H.; Juhasz, P.; Falick, A. M.; Vestal, M. L.; Martin, S. A. *J. Protein Chem.* **1997**, *16*, 363.
(109) Bruce, J. E.; Anderson, G. A.; Wen J.; Harkewitz, R.; Smith, R. D. *Anal. Chem.* **1999**, *71*, 2595.
(110) Goodlett, D. R.; Bruce, J. E.; Anderson, G. A.; Rist, B.; Pasa-Tolic, L.; Fiehn, O.; Smith, R. D.; Aebersold, R. *Anal. Chem.* **2000**, *72*, 1112.
(111) Pappin, D. J. C.; Rahman, D.; Hansen, H. F.; Bartlet-Jones, M.; Jeffery, W.; Bleasby, A. J. *Mass Spectrometry in the Biological Sciences*; Humana Press: Totowa, 1995; p 135.
(112) Craig, A. G.; Fischer, W. H.; Rivier, J. E.; McIntosh, J. M.; Gray, W. R. *Techniques in Protein Chemistry VI*; Academic Press: San Diego, 1990; p 31.
(113) Sechi, S.; Chait, B. T. *Anal. Chem.* **1998**, *70*, 5150.
(114) James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Protein Sci.* **1994**, *3*, 1347.
(115) Jensen, O. N.; Vorm, O.; Mann, M. *Electrophoresis* **1996**, *17*, 938.
(116) Woods, A. S.; Huang, A. Y. C.; Cotter, R. J.; Pasternack, G. R.; Pardoll, D. M.; Jaffee, E. M. *Anal. Biochem.* **1995**, *226*, 15.
(117) Patterson, S. D. *Electrophoresis* **1995**, *16*, 1104.
(118) Wiley: W. C.; McLaren, I. H. *Rev. Sci. Instrum.* **1953**, *26*, 1150.
(119) Vestal, M. L.; Juhasz, P.; Martin, S. A. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1044.
(120) Brown, R. S.; Lennon, J. J. *Anal. Chem.* **1995**, *67*, 1998.
(121) Urquhart, B. L.; Atsalos, T. E.; Roach, D.; Basseal, D. J.; Bjellqvist, B.; Britton, W. L.; Humphery-Smith, I. *Electrophoresis* **1997**, *18*, 1384.
(122) Zubarev, R. A.; Hakansson, P.; Sundqvist, B. *Anal. Chem.* **1996**, *68*, 4060.
(123) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Anal. Chem.* **1991**, *63*, 2488.
(124) Light-Wahl, K. J.; Loo, J. A.; Edmonds, C. G.; Smith, R. D.; Witkowska, H. E.; Shackleton, C. H.; Wu, C. S. *Biol. Mass Spectrom.* **1993**, *22*, 112.
(125) Lennon, J. J.; Walsh KA *Protein Sci.* **1997**, *6*, 2446.
(126) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 563.
(127) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III. *Nat. Biotechnol.* **1999**, *17*, 676.
(128) Suckau, D.; Cornett, D. S. *Analusis* **1998**, *26*, M18.
(129) Biemann, K. *Methods Enzymol.* **1990**, *193*, 886.
(130) Roepstorff, P.; Fohlman, J. *J. Biomed. Mass Spectrom.* **1984**, *11*, 601.
(131) Harrison, A. G.; Csizmadia, I. G.; Tang, T. H. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 427.
(132) Loo, J. A.; Edmonds, C. G. Smith, R. D. *Anal. Chem.* **1993**, *65*, 425.
(133) Schwartz, B. L.; Bursey, M. M. *Biol. Mass Spectrom.* **1992**, *21*, 92.
(134) Goodlett, D. R.; Gale, D. C.; Guiles, S.; Crowther, J. *Encyclopedia of Analytical Chemistry*; John Wiley: New York, 2000.
(135) Mak, M.; Mezo, G.; Skribanek, Z.; Hudecz, F. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 837.
(136) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. *J. Am. Chem. Soc.* **1996**, *118*, 8365.
(137) Spengler, B.; Luetzenkirchen, F.; Metzger, S.; Chaurand, P.; Kaufmann, R.; Jeffery, W.; Bartlet-Jones, M.; Pappin, D. J. C. *Int. J. Mass Spectrom. Ion Processes* **1997**, *169/170*, 127.
(138) Griffiths, W. J. *EXS* **2000**, *88*, 69.
(139) Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. *Nature* **1996**, *379*, 466.
(140) Wilm, M.; Neubauer, G.; Mann, M. *Anal. Chem.* **1996**, *68*, 527.
(141) Figeys, D.; Corthals, G. L.; Gallis, B.; Goodlett, D. R.; Ducret, A.; Corson, M. A.; Aebersold, R. *Anal. Chem.* **1999**, *71*, 2279.
(142) Stahl, D. C.; Swiderek, K. M.; Davis, M. T.; Lee, T. D. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 532.
(143) Shabanowitz, J.; Settlage, R. E.; Marto, J. A.; Christian, R. E.; White, F. M.; Russo, P. S. W.; Martin, S. E.; Hunt, D. F. *Mass Spectrometry in Biology and Medicine*; Human Press: Totowa, 2000; p 163.
(144) Ducret, A.; van Oostveen, I.; Eng, J. K.; Yates, J. R., III; Aebersold, R. *Protein Sci.* **1998**, *7*, 706.
(145) Courchesne, P. L.; Jones, M. D.; Robinson, J. R.; Spahr, C. S.; McCracken, S.; Bentley, D. L.; Luethy, R.; Patterson, S. D. *Electrophoresis* **1998**, *19*, 956.
(146) Goodlett, D. R.; Wahl, J. H.; Udseth, H. R.; Smith, R. D. *J. Microcolumn Sep.* **1993**, *5*, 57.
(147) Susin, S. A.; Lorenzo, H. K.; Zamzmi, N.; Marzo, I.; Snow, B. E.; Brothers, G. M.; Mangion, J.; Jacotot, E.; Costantini, P.; Loeffler, M. M.; Larochette, N.; Goodlett, D. R.; Aebersold, R.; Siderovski, D. P.; Penninger, J. M.; Kroemer, G. *Nature* **1999**, *397*, 441.
(148) McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R., III. *Anal. Chem.* **1997**, *69*, 767.
(149) Chait, B. T.; Wang, R.; Beavis, R. C.; Kent, S. B. H. *Science* **1993**, *262*, 89.
(150) Wang, R.; Chait, B. T.; Kent, S. B. H. In *Techniques in Protein Chemistry V*; Academic Press: San Diego, 1994; p 19.
(151) Bartlet-Jones, M.; Jeffery, W. A.; Hansen, H. F.; Pappin, D. J. C. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 737.
(152) Yates, J. R., III; McCormack, A. L.; Hayden, J. B.; Davey, M. P. *Cell Biology: A Laboratory Handbook*; Academic Press: New York, 1994; p 380.
(153) Takao, T.; Hori, H.; Okamoto, K.; Harada, A.; Kamachi, M.; Shimonishi, Y. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 312.
(154) Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. *Electrophoresis* **1996**, *17*, 945.
(155) Kosaka, T.; Takazawa, T.; Nakamura, T. *Anal. Chem.* **2000**, *72*, 1179.
(156) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015.
(157) Qin, J.; Herring, C. J.; Zhang, X. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 209.
(158) Huang, Z.-H.; Shen, T.; Wu, J.; Gage, D. A.; Watson, J. T. *Anal. Biochem.* **1999**, *268*, 305.
(159) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 7131.
(160) Yates, J. R., III. *Trends Genet.* **2000**, *16*, 5.
(161) Mortz, E.; O'Connor, P. B.; Roepstorff, P.; Kelleher, N. L.; Wood, T. D.; McLaffertey, F. W.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8264.
(162) Mann, M.; Wilm, M. *Anal. Chem.* **1994**, *66*, 4390.
(163) Hoving, S.; Munchbach, M.; Schmid, H.; Signor, L.; Lehmann, A.; Staudenmann, W.; Quadroni, M.; James, P. *Anal. Chem* **2000**, *72*, 1006.
(164) Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. *J. Comput. Biol.* **1999**, *6*, 327.
(165) Eng, J. K.; McCormack, A. L.; Yates, J. R., III. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976.
(166) Cox, A. L.; Skipper, J.; Chen, Y.; Henderson, R. A.; Darrow, T. L.; Shabanowitz, J.; Engelhard, V. H.; Hunt, D. F.; Slingluff, C. L. *Science* **1994**, *264*, 716.
(167) Yates, J. R., III; Eng, J. K.; McCormack, A. L.; Schieltz, D. *Anal. Chem.* **1995**, *67*, 1426.
(168) Yates, J. R., III; Eng, J. K.; McCormack, A. L. *Anal. Chem.* **1995**, *67*, 3202.
(169) Bruce, J. E.; Anderson, G. A.; Brands, M. D.; Pasa-Tolic, L.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 416.
(170) Solouki, T.; Marto, J. A.; White, F. M.; Guan, S.; Marshall, A. G. *Anal. Chem.* **1995**, *67*, 4139.
(171) Goodlett, D. R.; Bruce, J. E.; Smith, R. D.; Aebersold, R. Unpublished data from collaboration at B. M. I. Environmental and Molecular Sciences Laboratory in Richland, WA, 1998.

(172) O'Farrell, P. H. *J. Biol. Chem.* **1975**, *250*, 4007.
(173) Klose, J. *Humangentik* **1975**, *26*, 231.
(174) Gygi, S. P.; Rochon, Y.; Franza, B.; Aebersold, R. *Mol. Cell. Biol.* **1999**, *19*, 1729.
(175) Kurland, C. G. *FEBS Lett* **1991**, *285*, 165.
(176) Futcher, B. *Methods Cell Sci.* **1999**, *21*, 79.
(177) Perrot, M.; Sagliocco, F.; Mini, T.; Monribot, C.; Schneider, U.; Shevchenko, A.; Mann, M.; Jeno, P.; Boucherie, H. *Electrophoresis* **1999**, *20*, 2280.
(178) Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C. *Electrophoresis* **2000**, *21*, 1104.
(179) Washburn, M. P.; Yates, J. R. *Curr. Opin. Microbiol.* **2000**, *3*, 292.
(180) Tong, W.; Link, A.; Eng, J. K.; Yates, J. R., III. *Anal. Chem.* **1999**, *71*, 2270.
(181) Jensen, P. K.; Pasa-Tolic, L.; Peden, K. K.; Martinovic, S.; Lipton, M. S.; Anderson, G. A.; Tolic, N.; Wong, K. K.; Smith, R. D. *Electrophoresis* **2000**, *21*, 1372.
(182) Patterson, S. D. *Anal. Biochem.* **1994**, *221*, 1.
(183) Opiteck, G. J.; Jorgenson, J. W.; Anderegg, R. J. *Anal. Chem.* **1997**, *69*, 2283.
(184) Opiteck, G. J.; Lewis, K. C.; Jorgenson, J. W.; Anderegg, R. J. *Anal. Chem.* **1997**, *69*, 1518.
(185) Hsieh, Y. L.; Wang, H.; Elicone, C.; Mark, J.; Martin, S. A.; Regnier, F. *Anal. Chem.* **1996**, *68*, 455.
(186) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994.
(187) Kaufmann, R.; Kirsch, D.; Spengler, B. *Int. J. Mass Spectrom. Ion Processes* **1994**, *131*, 355.
(188) Masselon, C.; Anderson, G. A.; Harkewicz, R.; Bruce, J. E.; Pasa-Tolic, L.; Smith, R. D. *Anal. Chem.* **2000**, *72*, 1918.
(189) DeRisis, L.; Iyer, V. R.; Brown, P. O. *Science* **1997**, *278*, 680.
(190) De Leenheer, A. P.; Thienpont, L. M. *Mass Spectrom. Rev.* **1992**, *11*, 249.
(191) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6591.
(192) Pasa-Tolic, L.; Jensen, P. K.; Anderson, G. A.; Lipton, M. S.; Peden, K. K.; Martinovic, S.; Tolic, N.; Bruce, J. E.; Smith, R. D. *J. Am. Chem. Soc.* **1999**, *121*, 7949.
(193) Selle, H.; Schrader, M.; Schoeffski, P.; Hess, R.; Zucht, H.-D.; Heine, G.; Juergens, M.; Schulz-Knappe, P. Presented at the European Meeting on Biomarkers of Organ Damage and Dysfunction, Cambridge, U.K., April 3−7, 2000.
(194) Shrader, M. BioVisioN GmbH, Hannover, Germany, personal communication, 2000.
(195) Liu, Y.; Patricelli, M. P.; Cravatt, B. F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14694.
(196) Simone, N. L.; Bonner, R. F.; Gillespie, J. W.; Emmert-Buck, M. R.; Liotta, L. A. *Trends Genet.* **1998**, *14*, 272.
(197) Simone, N. L.; Remaley, A. T.; Charboneau, L.; Petricoin, E. F., III; Glickman, J. W.; Emmert-Buck, M. R.; Fleisher, T. A.; Liotta, L. A. *Am. J. Pathol.* **2000**, *156*, 445.
(198) Lee, H.; Williams, S. K.; Giddings, J. C. *Anal. Chem.* **1998**, *70*, 2495.
(199) Rout, M. P.; Aitchison, J. D.; Suprapto, A.; Hjertaas, K.; Zhao, Y.; Chait, B. T. *J. Cell Biol.* **2000**, *148*, 635.
(200) Heller, M.; Goodlett, D. R.; Watts, J. D.; Aebersold, R. *Electrophoresis* **2000**, *21*, 2180.
(201) Bouveret, E.; Rigaut, G.; Shevchenko, A.; Wilm, M.; Seraphin B. *EMBO J.* **2000**, *19*, 1661.
(202) Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; Seraphin, B. *Nat. Biotechnol.* **1999**, *17*, 1030.
(203) Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T. A.; Judson, R. S.; Knight, J. R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; Qureshi-Emili, A.; Li, Y.; Godwin, B.; Conover, D.; Kalbfleisch, T.; Vijayadamodar, G.; Yang, M.; Johnston, M.; Fields, S.; Rothberg, J. M. *Nature* **2000**, *403*, 623.
(204) Krishna, R. G.; Wold, F. *PROTEINS: Analysis & Design*; Academic Press: San Diego, 1998; p 121.
(205) Charbonneau, H.; Tonks, N. K. *Annu. Rev. Cell Biol.* **1992**, *8*, 463.
(206) Fischer, E. H.; Krebs, E. G. *Biochim. Biophys. Acta* **1989**, *1000*, 297.
(207) Hunter, T. *Cell* **1987**, *50*, 823.
(208) Duclos, B.; Marcandier, S.; Cozzone, A. J. *Methods Enzymol.* **1991**, *201*, 10.
(209) Karin, M.; Hunter, T. *Curr. Biol.* **1995**, *5*, 747.
(210) Johnson, L. N.; Barford, D. *J. Biol. Chem.* **1996**, *265*, 2409.
(211) Joung, I.; Kim, T.; Stolz, L. A.; Payne, G.; Winkler, D. G.; Walsh, C. T.; Strominger, J. L.; Shin, J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5778.
(212) Watts, J. D.; Welham, M. J.; Kalt, L.; Schrader, J. W.; Aebersold, R. *J. Immunol.* **1993**, *151*, 6862.
(213) Pawson, T.; Nash, P. *Genes Dev.* **2000**, *14*, 1027.
(214) Wange, R. L.; Isakov, N.; Burke, T. R.; Otaka, A.; Roller, P. P.; Watts, J. D.; Aebersold, R.; Samelson, L. E. *J. Biol. Chem.* **1995**, *270*, 944.
(215) Watts, J. D.; Affolter, M.; Krebs, D. L.; Wange, R. L.; Samelson, L. E.; Aebersold, R. *J. Biol. Chem.* **1994**, *269*, 29520.
(216) Watts, J. D.; Brabb, T.; Bures, E. J.; Wange, R. L.; Samelson, L. E.; Aebersold, R. *FEBS Lett.* **1996**, *398*, 217.
(217) Haspel, R. L.; Darnell, J. E., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10188.
(218) Abu-Amer, Y.; Ross, F. P.; McHugh, K. P.; Livolsi, A.; Peyron, J. F.; Teitelbaum, S. L. *J. Biol. Chem.* **1998**, *273*, 29417.
(219) Garrison, T. R.; Zhang, Y.; Pausch, M.; Apanovitch, D.; Aebersold, R.; Dohlman, H. G. *J. Biol. Chem.* **1999**, *274*, 36387.
(220) Boyle, W. J.; van der Geer, P.; Hunter, T. *Methods Enzymol.* **1991**, *201*, 110.
(221) Martzen, M. R.; McCraith, S. M.; Spinelli, S. L.; Torres, F. M.; Fields, S.; Grayhack, E. J.; Phizicky, E. M. *Science* **1999**, *286*, 1153.
(222) Zhou, H.; Watts, J. D.; Aebersold, R. Presented at the 48th American Society for Mass Spectrometry Meeting, Long Beach, CA, June 2000.
(223) Weckwerth, W.; Willmitzer, L.; Fiehn, O. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1677.
(224) Gallis, B.; Corthals, G. L.; Goodlett, D. R.; Ueba, H.; Kim, F.; Presnell, S. R.; Figeys, D.; Harrison, D. G.; Berk, B. C.; Aebersold, R.; Corson, M. *J. Biol. Chem.* **1999**, *274*, 30101.
(225) Watts, J. D.; Krebs, D. L.; Wange, R. L.; Samelson, L. E.; Aebersold, R. *Biochemical and Biotechnological Applications of Electrospray Ionization Mass Spectrometry*; American Chemical Society: Washington, DC, 1996; p 381.
(226) Yanagida, M.; Miura, Y.; Yagasaki, K.; Taoka, M.; Isobe, T.; Takahashi, N. *Electrophoresis* **2000**, *21*, 1890.
(227) Heffetz, D.; Fridkin, M.; Zick, Y. *Eur. J. Biochem.* **1989**, *182*, 343.
(228) Wu, J. J.; Yarwood, D. R.; Pham, Q.; Sills, M. A. *J. Biomol. Screening* **2000**, *5*, 23.
(229) Soskic, V.; Gorlach, M.; Poznanovic, S.; Boehmer, F. D.; Godovac-Zimmerman, J. *Biochemistry* **1999**, *38*, 1757.
(230) Goodlett, D. R.; Aebersold, R.; Watts, J. D *Rapid Commun. Mass Spectrom.* **2000**, *14*, 344.
(231) Verma, R.; Annan, R. S.; Huddleston, M. J.; Carr, S. A.; Reynard, G.; Deshaies, R. J. *Science* **1997**, *278*, 455.
(232) Gatti, A.; Traugh, J. A. *Anal. Biochem.* **1999**, *266*, 198.
(233) Porath, J. *Protein Expression Purif.* **1992**, *3*, 263.
(234) Boyle, W. J.; Smeal, T.; Defize, L. H.; Angel, P.; Woodgett, J. R.; Karin, M.; Hunter, T. *Cell* **1991**, *64*, 573.
(235) Affolter, M.; Watts, J. D.; Krebs, D. L.; Aebersold, R. *Anal. Biochem.* **1994**, *223*, 74.
(236) Neville, D. C.; Rozanas, C. R.; Price, E. M.; Gruis, D. B.; Verkman, A. S.; Townsend, R. R. *Protein Sci.* **1997**, *6*, 2436.
(237) Nuwaysir, L. M.; Stults, J. T. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 662.
(238) Tempst, P.; Link, A. J.; Riviere, L. R.; Fleming, M.; Elicone, C. *Electrophoresis* **1990**, *11*, 537.
(239) Katze, M. G.; Kwieciszewski, B.; Goodlett, D. R.; Blakely, C. M.; Nedderman, P.; Tan, S.-L.; Aebersold, R. *Virology* **2000**, *278*, 501.
(240) Annan R. S.; Carr S. A. *J. Protein Chem.* **1997**, *16*, 391.
(241) Covey, T. R.; Huang, E. C.; Henion, J. D. *Anal. Chem.* **1991**, *63*, 1193.
(242) Katta, V.; Chowdhury, S. K.; Chait, B. T. *Anal. Chem.* **1991**, *63*, 174.
(243) Cao, P.; Stults J. T. *J. Chromatogr.* A **1999**, *853*, 225.
(244) Meyer, H. E.; Eisermann, B.; Heber, M.; Hoffmann-Posorske, E.; Korte, H.; Weigt, C.; Wegner, A.; Hutton, T.; Donella-Deana, A.; Perich, J. W. *FASEB J.* **1993**, *7*, 776.
(245) Huddleston, M. J.; Annan, R. S.; Bean, M. F.; Carr, S. A. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 710.
(246) Hunter, A. P.; Games, D. E. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 559.
(247) Aebersold, R.; Figeys, D.; Gygi, S.; Corthals, G.; Haynes, P.; Rist, B.; Sherman, J.; Shang, Y.; Goodlett, D. *J. Protein Chem.* **1998**, *17*, 533.
(248) Gibson, B. W.; Cohen, P. *Methods Enzymol.* **1990**, *193*, 480.
(249) Carr S. A.; Huddleston M. J.; Annan, R. S. *Anal. Biochem.* **1996**, *239*, 180.
(250) Neubauer, G.; Mann, M. *Anal. Chem.* **1999**, *71*, 235.
(251) Annan, R. S.; Carr, S. A. *Anal. Chem.* **1996**, *68*, 3413.
(252) DeGnore, J. P.; Qin, J. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 1175.
(253) Zhang, X.; Herring, C. J.; Romano, P. R.; Szczepanowska, J.; Brzeska, H.; Hinnebusch, A. G.; Qin, J. *Anal. Chem.* **1998**, *70*, 2050.
(254) Amankwa, L. N.; Harder, K.; Jirik, F.; Aebersold, R. *Protein Sci.* **1995**, *4*, 113.

CR990076H