Accurate Mass Measurements in Proteomics

Tao Liu, Mikhail E. Belov, Navdeep Jaitly, Wei-Jun Qian, and Richard D. Smith*

Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99354

Received November 1, 2006

Contents

1. Introduction 3621
   1.1. MS Based Proteomics Strategies 3621
   1.2. The Need for Accurate Mass Measurements 3623
2. Mass Measurement Accuracy 3623
   2.1. FTICR Mass Spectrometry 3624
      2.1.1. External Mass Calibration 3625
      2.1.2. Internal Mass Calibration 3626
   2.2. Orbitrap Mass Spectrometry 3628
   2.3. TOF Mass Spectrometry 3629
      2.3.1. MALDI-TOF 3629
      2.3.2. ESI-TOF 3630
3. Accurate Mass Measurements in Proteomics 3631
   3.1. Peptide Mass Fingerprinting 3631
   3.2. LC-MS/MS Analysis of Peptide Mixtures 3633
      3.2.1. Increased Confidence in Peptide Identification 3633
      3.2.2. *De Novo* Peptide Sequencing 3635
      3.2.3. Characterization of Post-translational Modifications 3636
   3.3. LC-MS Analysis of Peptide Mixtures 3638
      3.3.1. LC-MS Feature Based Profiling for High-Throughput Proteomics 3639
      3.3.2. LC-MS Feature Based Quantitative Proteomics 3640
   3.4. Intact Protein Analysis 3641
      3.4.1. Intact Protein Profiling 3641
      3.4.2. Protein Fragmentation and Characterization 3642
4. Informatics Algorithms and Pipelines for Interpreting and Applying Accurate Mass Information 3644
   4.1. Analysis Algorithms 3644
   4.2. Analysis Pipelines 3645
5. Conclusions and Outlook 3647
6. Abbreviations 3647
7. Acknowledgments 3648
8. References 3648

1. Introduction

The ability to broadly identify and measure abundances for biological macromolecules, especially proteins, is essential for delineating complex cellular networks and pathways in systems biology studies. Enabled by the development in the late 1980s of two "soft" ionization methods—electrospray ionization (ESI)[1] and matrix-assisted laser desorption/ionization (MALDI)[2,3] that prevent or limit fragmentation of large biomolecules—and the increasing availability of genomic sequence databases, mass spectrometry (MS) has become a major analytical tool for studying the array of proteins in an organism, tissue, or cell at a given time, i.e., for *proteomics*. Such proteome-wide analysis provides a wealth of biological information, such as sequence, quantity, post-translational modifications (PTMs), interactions, activities, subcellular distributions, and structure of proteins, that is critical to the comprehensive understanding of a biological system.

MS instrumentation and bioinformatics tools have rapidly evolved in recent years as a result of the ever increasing demands for more powerful analytical capabilities in protein biochemistry and the emerging field of systems biology. New types of mass analyzers and complex multistage and hybrid instruments provide new opportunities for diverse protein and proteome analyses.[4,5] In particular, instruments that afford accurate mass measurements are being increasingly applied in proteomics studies not only to determine protein identity but also to help determine protein PTM states, as well as interactions between proteins and other molecules in a more unambiguous and higher-throughput fashion than before.

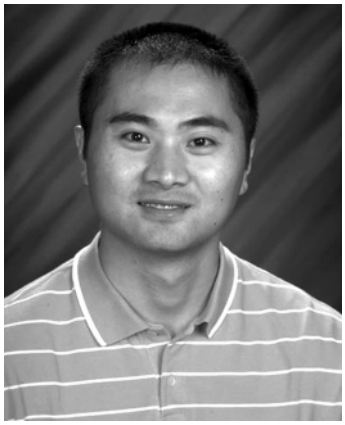
Herein, we review the presently most important and promising topics in proteomics applying accurate mass measurements rather than the broader area of proteomics, which has been discussed and summarized in many excellent reviews.[6−16] The two general approaches to MS based proteomics and a brief discussion on the need for accurate mass measurements complete this introduction prior to reviewing high-resolution MS instrumentation and methods that provide high mass measurement accuracy (MMA), improvements in proteomics applications applying accurate mass measurements, and developments in bioinformatics that utilize high-mass-accuracy data to enable new data analysis strategies.

1.1. MS Based Proteomics Strategies

In general, there are two different strategies for proteome analysis using MS. One strategy is the so-called "bottom-up" strategy [typically implemented as "shotgun" proteomics[17] or two-dimensional gel electrophoresis (2-DE)[18−20] coupled to peptide mass fingerprinting (PMF)[21−26]], which involves the conversion of proteins to peptides through either enzymatic digestion or chemical cleavage prior to MS analysis. Proteins can then be identified from mass measurements of a set of peptides derived from the parent protein (e.g., PMF) or from fragmentation of one or more of these

* Address correspondence to: Dr. Richard D. Smith, Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, P.O. Box 999, MSIN: K8-98, Richland, WA 99354 (rds@pnl.gov).

10.1021/cr068288j CCC: $65.00
Published on Web 07/25/2007



Tao Liu received his B.S. degree in Chemistry from Nanchang University, China, in 1996 and a Ph.D. degree in Biochemistry and Molecular Biology in 2001 from Shanghai Institute of Biochemistry, Chinese Academy of Sciences. He was a postdoctoral research associate at Howard Hughes Medical Institute at the University of Washington, Seattle. In 2003, he joined Pacific Northwest National Laboratory in Richland, WA, as a postdoctoral research fellow (2003–2005), and he remained at PNNL as a Senior Research Scientist (2005 to the present) in the Biological Sciences Division. His research interests include quantitative proteomics, protein post-translational modifications, and biomarker discovery and verification using mass spectrometry.

Mikhail Belov received his M.S. degree in Physics from Moscow Engineering Physics Institute, Russia, and his Ph.D. degree in Physics from General Physics Institute, Moscow, Russia. He was a Research Fellow at the University of Warwick, U.K., and a Senior Research Scientist at Pacific Northwest National Laboratory, Richland, WA. He then worked for over 3 years as a Principal Scientist at the start-up biotech company Predicant Biosciences, South San Francisco, CA. He is currently a Staff Scientist at Pacific Northwest National Laboratory. His research interests include gas/condensed phase separations and mass spectrometry of biomolecules. Dr. Belov is a coauthor on more than 40 refereed publications and a co-inventor of 7 patents. In 2003, he received an R&D 100 Award for the "Proteome Express" system.

peptides [using tandem MS (MS/MS)].[27−30] As a result of rapid developments in MS instrumentation that have increased speed and sensitivity and in database searching algorithms (e.g., SEQUEST and MASCOT),[31−37] these two MS based approaches quickly replaced the traditional Edman degradation approach[38] as the method of choice for protein identification.

The second strategy approaches proteome characterization from the "top-down"; i.e., individual proteins are selected for mass measurement of the whole protein, gas-phase fragmentation of the protein ions, and direct database searching.[39] While the top-down strategy is potentially capable of providing full sequence coverage and important information that might be unobtainable at the peptide level,

Navdeep Jaitly received his M.S. degree in Computer Science from the University of Waterloo. He worked as a software developer in IBM Toronto Labs and as a Senior Research Scientist and Group Leader in Bioinformatics at Caprion Pharmaceuticals in Montreal. Currently he is a Senior Research Scientist at Pacific Northwest National Laboratory. His research interests include the application of Machine Learning and Statistical techniques to analysis of proteomics data from mass spectrometry.

Wei-Jun Qian received his B.S. degree in Chemistry at Nanjing University, China, in 1994 and a Ph.D. degree in Bioanalytical Chemistry in 2002 from the University of Florida under the direction of Robert T. Kennedy. He joined Pacific Northwest National Laboratory following his graduation as a postdoctoral research fellow, where he is presently a Senior Research Scientist in the Biological Sciences Division. Dr. Qian's current research focuses on developing integrated mass spectrometry based approaches that enable quantitative measurements of the dynamics of proteins and protein modifications in biological and clinical applications.

e.g., protein point mutation, protein PTMs, and protein isoforms, all of which may be key factors that contribute to protein functions, the current top-down approaches are generally limited by throughput, separation peak capacity, and fragmentation efficiency that are typically inferior to those of the bottom-up methods. Protein sequence information can be obtained by using, for example, Fourier transform ion cyclotron resonance (FTICR) mass spectrometers along with fragmentation techniques, such as electron capture dissociation (ECD)[40] and collision-induced dissociation (CID). Top-down protein characterization can also be carried out by using proton-transfer reactions on ion trap (IT) instruments[41] or electron-transfer dissociation (ETD) on orbitrap mass spectrometers.[42] Both ECD and ETD have the advantage of providing complementary fragmentation of both peptide and proteins, thus greatly enhancing database searching for protein identification. Moreover, they allow labile PTMs such as phosphorylation to be retained, which in turn often allows unambiguous determination of modification

Richard D. Smith received his B.S. degree in Chemistry in 1971 from University of Massachusetts at Lowell and a Ph.D. degree in Physical Chemistry in 1975 from the University of Utah. Dr. Smith is a Battelle Fellow and Chief Scientist in the Biological Sciences Division at Pacific Northwest National Laboratory in Richland, WA. His research has involved the development and application of advanced methods and instrumentation and their applications in biological research and, particularly, proteomics. Dr. Smith is Director of the NIH Biomedical Technology Resource Center for Integrative Biology, the NIAID Biodefense Proteomics Research Center for Identifying Targets for Therapeutic Interventions using Proteomics, and the U.S. Department of Energy High Throughput Proteomics Facility at PNNL. He is an adjunct faculty member of the Departments of Chemistry at Washington State University, the University of Utah, and the University of Idaho. Dr. Smith has presented more than 350 invited or plenary lectures at national and international scientific meetings, and he is the author or coauthor of more than 600 publications. Dr. Smith holds 29 patents and has been the recipient of seven R&D 100 Awards.

sites. Therefore, by combining bottom-up and top-down strategies, a proteome or subset(s) of a proteome (e.g., phosphoproteome) can be studied in unprecedented detail.

1.2. The Need for Accurate Mass Measurements

There are significant challenges in proteomics analysis that stem from the tremendous complexity of biological systems and the range of protein abundances in systems of interest (often referred to as the "dynamic range" challenge). An example of an extreme case is the blood serum/plasma proteome, in which almost all expressed proteins can potentially be present and span a concentration range of at least 10 orders of magnitude, which exceeds the dynamic range of any present single MS based analytical method or instrument.[43] When proteins are converted to peptides by enzymatic cleavage, this already striking sample complexity is further increased. The presence of multiple protein forms (e.g., isoforms, post-translational modifications, and truncated forms that result from proteolysis) poses additional challenges for proteome analysis. A practical solution for addressing these issues is to use a "divide and conquer" sample fractionation strategy; for example, selectively analyzing subsets of the proteome that have been enriched by using different techniques.[44−47] Another fractionation strategy is to combine a high-efficiency separation such as high-resolution 2-DE or multiple dimension liquid chromatography[48,49] with MS. The use of different separation techniques in protein and peptide profiling can also provide very useful physical and chemical property information, e.g., molecular weight (M_r), isoelectric point (p*I*), hydrophobicity, and affinity to certain matrices, that is useful for improving protein identifications.

Regardless of the level of separation, identification of peptides/proteins by either MS or MS/MS typically relies on matching parent ions or fragment ion masses to a theoretical database derived from protein sequences for a given genome. The confidence of identifications strongly depends on the accuracy of the mass measurements, especially in the case of highly complex samples derived from higher organisms (e.g., human). It is well-known that the number of possible amino acid composition candidates rapidly decreases with increasing MMA.[35,50−55] For example, a MMA of ±1 part per million (ppm) can exclude 99% of peptides that have the same nominal mass but different elemental and amino acid compositions, which results in a high degree of confidence in peptide characterization.[53] One of the most popular types of tandem mass spectrometers that are being used in proteomics studies, the linear ion traps, are capable of acquiring hundreds to tens of thousands of tandem mass spectra over the course of one liquid chromatography separation (LC-MS/MS); however, the MMA achievable is generally low.[56] Thus, a large percentage of proteins can be misidentified, depending on the scoring criteria used to "filter" MS/MS data that are searched against a database.[57−60] The use of high scoring thresholds can significantly lower the false discovery rate (FDR), but at the expense of losing a fraction of the true positive peptide identifications. Various statistical approaches have been developed to estimate the FDR in a given data set to ensure that quality protein identifications can be made through large scale MS/MS experiments;[57,58,60−65] however, obtaining confident peptide identification remains challenging with these approaches.

The specificity of peptide identifications can be significantly improved by using multiple MS stages (MS^n)[66,67] or complementary fragmentation techniques (e.g., ECD combined with CID[68]), as well as by measuring the mass of peptide ions at high MMA.[35,51,53,69,70] Although MS/MS analysis is effective for identifying peptides and proteins, the number of detectable peptides that elute during a typical LC-MS/MS analysis generally far exceeds the ability of the tandem mass spectrometer to perform CID on all of them: "too many peptides; too little time". In addition, a comprehensive proteome analysis often requires information regarding temporal changes in protein expression be collected on a global scale, which demands a high-throughput MS capability for in-depth and reproducible protein identification and quantification from substantially identical samples. These needs can be addressed by using the concept of an "accurate mass and time (AMT) tag"; that is, if the mass of a peptide can be measured with sufficient MMA along with accurately measured LC elution time such that the detected LC-MS feature is unique in the mass and time space among all possible peptide candidates in a mass and time tag database pre-established for the proteome using LC-MS/MS, then it can be used as an AMT tag for higher-throughput peptide/protein identification by circumventing the need for repetitive MS/MS measurements.[71]

2. Mass Measurement Accuracy

There is a general lack of a single clear definition of mass accuracy in the field of proteomics.[72] In the classical definition, accuracy is a degree of conformity of the measured (or calculated) quantity to its true value. Precision determines the degree to which measured (or calculated) quantities show the same or similar result. In biological mass spectrometry, one of the objectives is to accurately determine a mass-to-charge ratio (m/z) of the biomolecules of interest and, thereby, obtain their accurate masses using a "deiso-

toping" algorithm. However, mass spectrometers experimentally measure parameters other than *m*/*z* [i.e., reduced cyclotron frequency in FTICR MS or ion's arrival time at the detector in time-of-flight (TOF) MS], and calibration procedures are needed to convert the measured quantities to *m*/*z* values. Since experimentally measured parameters are often affected by the complexity of the studied system that represents an ensemble of many particles and by the nonideality of an experimental apparatus, sophisticated correction routines generally need to be introduced into calibration procedures to mitigate experimental imperfections if high accuracy is to be achieved. Among the most prominent factors affecting the accuracy of conversion of the experimentally measured quantities to *m*/*z*'s are the space charge effect, fringing field effects, detector and acquisition system dependence on the ion abundance, etc. Correction routines enable reduction of mass measurement errors to sub-ppm levels in a single measurement. Given multiple species are present in a mass spectrum, an average or root-mean-square (rms) mass measurement error is introduced as a metric of mass accuracy.

In a typical large-scale proteomic study, analyte detection is augmented by high-performance separation of biomolecules in the condensed phase, e.g., using on-line capillary liquid chromatography (LC) or capillary electrophoresis (CE) upstream of a mass spectrometer. This results in multiple measurements of the same analyte over its elution/migration profile from an LC/CE column and yields a distribution of mass measurement errors that implies the use of statistical tools. Based on the experimentally observed mass error distributions (Gaussian type, gamma distribution, etc.), several metrics that reflect the experimental accuracy on the global scale are introduced. Each observed feature is characterized by the mean error and the variance, and the whole dataset, that may include $>10^5$ features, is represented by the distribution of mean errors of individual features. A single metric that reflects the accuracy of measurement in a large-scale proteomic experiment is then represented by the width of the above statistical error distribution within the 95% confidence interval. A similar approach is used for the normalized retention times of the observed features. The features that fit within the predetermined range (for example, 2 variances) of mass measurement and retention time error distributions are searched against a genome database, yielding peptide identifications. The latter are subject to further statistical analysis aimed at establishing an FDR. Such an approach enables objective control of the measurement quality based on orthogonal characteristics such as MMA and analyte retention time, and the resulting peptide identifications are obtained with a well-defined FDR.

Mass calibration procedures employed with MS instrumentation can be separated to external and internal calibrations. External calibration employs a set of fixed calibration coefficients in the course of the entire proteomic experiment, often comprising hundreds of mass spectra. External calibration relies on the stability of instrumental parameters and may result in significant errors if some of the parameters are affected, for example, by temperature drift, space-charge fluctuations, timing jitters, etc. Internal calibration is based on mixing one or several standards or calibrants of known *m*/*z* values with the analyte and then deriving the *m*/*z* values of the unknown species from the calibration equation obtained with the standards. Though internal calibration is more robust to variations in instrumental parameters, some of the experimental deviations (e.g., excessive ion populations in the ICR traps) lead to nonlinear effects that reduce MMA.

High mass resolving power is required to achieve sufficient *precision* for accurate mass assignment. Though lower-resolution mass spectrometers can achieve high *accuracy*, their application is limited to the analysis of target compounds that are well-separated from other species in the *m*/*z* domain. For instance, triple quadrupole (TQ) instruments are best suited to operate in selected ion monitoring (a particular ion or set of ions is monitored) and selected/multiple reaction monitoring modes (parent ions of a certain type and their fragment ions are detected). These techniques are predominantly applied to the trace analysis of compounds that are well-characterized in previous studies. Global analysis of a complex sample with, e.g., TQ mass spectrometers operating with unit resolution in precursor ion scanning mode is limited to species that differ by more than ± 0.5 Da. The need for high-resolution instrumentation is further exacerbated in proteomic experiments and often represents a challenge for accurate and precise mass determination of isotopic distributions that are significantly different in ion abundances and are closely spaced (sometimes overlapped) in the *m*/*z* domain.

MS instrumentation capable of attaining low-ppm MMA and high resolving power in a typical proteomic experiment is presently limited to FTICR,[73] orthogonal TOF,[74,75] and recently developed 3D electrostatic ion trap (orbitrap) mass spectrometers.[76] Measurement specifics for each of these spectrometers follow.

2.1. FTICR Mass Spectrometry

Cyclotron motion was first employed in mass spectrometry in the late 1940s with the introduction of the first ICR mass spectrometer, called the omegatron.[77] In this first device, excitation was performed by applying a continuous field at the ion cyclotron frequency, which resulted in charge detection on a small collector blade. A mass spectrum was obtained by scanning the electromagnet field to bring ions of different *m*/*z* into resonance. Since its inception in 1973,[78] FTICR has been the subject of multiple reviews,[79−86] several journal issues,[87,88] and several books[89,90] that give a full-range technical introduction to ion cloud behavior in combined magnetic and electric fields, subsequent signal processing, and technique applications. The reader is referred to these publications for more information. The application of FTICR in proteomics has also been recently reviewed.[91,92]

FTICR is well-known for obtaining high mass resolution and has been experimentally demonstrated to exhibit a mass resolving power of ~8 000 000 in an analysis of bovine ubiquitin (8559.6 Da), which is sufficient to distinguish the isotopic fine structure of the protein.[93] This ultrahigh resolving power was obtained in a high magnetic field of 9.4-Tesla (T) at a reduced number of ions and an increased postexcitation radius. The number of trapped ions was then further reduced by applying the stored waveform inverse Fourier transform (SWIFT) ejection[94] of all charge states but one of a given protein. Following ejection of the unwanted species, electrostatic potentials on the end-cap electrodes of the trap were reduced to a few tenths of a volt over a minute-long period to allow for efficient translational "evaporative" cooling of the remaining ion ensemble. In a typical proteomic experiment, the time scale for accurate mass measurement is limited to ~1 s. This time scale poses a constraint on the maximum achievable resolving power that is dependent on the *m*/*z* of the analyzed ions and typically limited to ~100 000.

Ultrahigh mass accuracy and precision are achievable with FTICR for several reasons.[95] First, mass is determined by measuring cyclotron frequency, a parameter measurable with extremely high precision. Second, superconducting magnets routinely achieve a time stability of a few parts per billion per hour (ppb/h), providing the time stability of the measurement. Third, the behavior of ions near the center of an ICR trap is very accurately described by a three-dimensional quadrupolar potential. Therefore, the frequency of ion axial oscillation is independent of the ion coordinate near the center of the trap. Fourth, the rapid cyclotron and axial motions of an ion effectively time average spatial nonidealities. At sufficiently long transients (~1−10 s), the slower magnetron motion incurred (e.g., in side-kick trapping[96]) is also time averaged. Given the low ion population in the ICR trap, both mass precision and accuracy have been shown to be in the sub-ppm range.[97] However, the precision of high-resolution FTICR does not guarantee the accuracy of measurement, as systematic effects can produce deviations between measured and calculated mass values.

To better understand factors that affect the detected cyclotron frequencies in an ICR trap, it is important to consider the FTICR detection system. An ion cloud trapped in a combined trap experiences four basic motions that include cyclotron motion, magnetron motion, axial oscillation, and rotation around its central axis.[98] The attraction between the space charge of an ion cloud and its image charge in the trap walls causes a slow drift around the trap's central axis, in addition to the magnetron drift caused by the trapping fields.[99,100] In conventional non-neutral plasma experiments, this image-induced drift is dominant and the motion it causes is called diocotron motion.[98] As a result, the detected cyclotron frequency, ω_{ICR}, is a superposition of the fast and slow oscillation frequencies in the trap:

$$\omega_{ICR} = \Omega - \omega_M - \omega_D - \delta_{sc} \quad (1)$$

$$\omega_M = \frac{\Omega}{2} - \frac{\Omega}{2}\sqrt{1 - \frac{2\omega_z^2}{\Omega^2}} \approx \frac{V_t}{2|B|d^2} \quad (2)$$

$$\omega_Z = \sqrt{\frac{azV_t}{md^2}} \quad (3)$$

$$\omega_D \approx \left(\frac{\rho_c^2}{r_w^2}\right)\omega_R \quad (4)$$

where ω_{ICR}, Ω, ω_M, and ω_D are the detected, unperturbed, magnetron, and diocotron frequencies, respectively; ω_z is the frequency of the axial oscillation; δ_{sc} is the space-charge term;[101] a is the geometry factor; V_t is the trapping voltage; B is the uniform magnetic field; d is the characteristic length of the trap; m/q is the mass-to-charge ratio of the ion; ρ_c and r_w are the ion cloud and trap wall radii,[102] respectively; and ω_R is the ion cloud rotation frequency due to $\mathbf{E} \times \mathbf{B}$ drift.

Equations 1−4 show that the detected cyclotron frequency depends on the axial oscillation frequency, the number of ions in the trap, and the ion cloud interaction with its image charge. Low-m/z ions also experience relativistic shifts in the measured cyclotron frequency;[103] the effect is typically ignored in experiments with higher-m/z ions detected, such as in proteomics. In theory, an ion postexcitation radius is independent of the m/z[78,104]

$$r = \frac{V_{p-p}T_{excite}}{2dB_0} \quad (5)$$

where V_{p-p} is the peak-to-peak voltage, T_{excite} is the excitation period, d is the distance between excite plates, and B is the magnetic field. However, in experiments, due to a nonideal spatial distribution of the excite field within an ICR trap, ions would have some narrow radial distribution that is broadened by the space charge. Any deviations of the axial field distribution from the ideal harmonic potential would then result in an axial oscillation frequency (and the measured cyclotron frequency) dependence on the ion radial position and lead to frequency shifts. As a result, ions positioned at the axial periphery of an ion cloud would be "evaporating" from the coherent ensemble, creating comet-like structures that were observed with supercomputer modeling.[105] An increase in the total number of trapped ions would result in further elongation of an ion cloud along the trap axis and pushing of the ion cloud into the trap regions with inharmonic field distribution, thus further exacerbating frequency shifts.

Another source of frequency shifts results from the interaction of ion clouds in the ICR trap. Using a simplified model of two Coulombically interacting ion clouds, both positive and negative frequency shifts have been predicted for the point charge model and then verified by numerical simulations.[106] In particular, the numerical simulations revealed that a spherical ion cloud with a cyclotron radius smaller than a second spherical ion cloud experiences a positive frequency shift from the second ion cloud, contrary to the negative frequency shifts caused by the total space charge as described by eq 1. These "local" frequency shifts have practical implications for FTICR mass calibration at a MMA of better than 1 ppm.

2.1.1. External Mass Calibration

The theoretical framework of space-charge-induced frequency shifts[100] has been used to develop an expression that relates observed frequencies, ω_{obs}, to m/z:[107]

$$\omega_{obs} = \frac{qB}{m} - \frac{2\alpha V_t}{d^2 B} - \frac{q\rho G_i}{\epsilon_0 B} \quad (6)$$

The last term represents the space-charge component of the mass shift, where ρ is the ion cloud density and G_i is the ion cloud geometry. Sub-ppm mass accuracy was demonstrated using this relationship for low-m/z ions by correlating the shift between the internal reference mass and the measured mass.[107] Parametrization of the mass-frequency relationship yielded an equation which is widespread for FTICR mass calibration:[84,108]

$$\frac{m}{z} = \frac{a}{f} + \frac{b}{f^2} \quad (7)$$

where a and b are the parameters determined in the experiment. The second-order frequency term accounts for the shifts that arise from applied and induced electric fields. Although the space-charge term is included, variations in ion populations severely degrade the ability of this equation to predict frequencies for externally calibrated reference masses, as b is a function of the density of the ions used to calibrate the mass spectrum.

In 4.7-T FTICR experiments with matrix-assisted laser desorption/ionization of high-molecular-weight polymers

with a wide mass distribution, mass errors of 100 ppm or more were reported for externally calibrated mass spectra when ion intensities were not taken into account. By matching the total ion intensities of calibrant and analyte mass spectra, the protonated ion of the insulin B-chain (3494.6513 Da) was measured with high accuracy (average of 10 measurements, $\sigma = 2.3$ ppm, average absolute error 1.6 ppm) using a polymer sample as an external calibrant.[109] A calibration equation with a higher-order correction term was proposed,

$$\frac{m}{z} = \frac{A}{f} + \frac{B}{f^2} + \frac{C}{f^3} \tag{8}$$

with a caveat that the calibration constants *A*, *B*, and *C* would be accurate only for the mass spectra that have the same total intensity as that of the calibrant mass spectrum from which the constants were derived. Given a mass spectrum with arbitrary ion intensities, linear interpolation of the frequencies that would have been measured if the total ion intensities were the same resulted in an ion-number-corrected calibration equation, where the experimental frequency, *f*, in eq 8 was replaced by the estimated frequency, $f_{\text{estimated}}$:

$$f_{\text{estimated}} = f_{\text{measured}} + c(I_{\text{calibrant}} - I_{\text{analyte}}) \tag{9}$$

Following this correction, a mass accuracy of 2.0 ppm (average of 20 measurements, $\sigma = 4.2$ ppm, average absolute error of 3.5 ppm) was achieved. It is important to note that the highest linearity in frequency versus intensity for MALDI-generated ions was obtained by using suspended trapping[110] with collisional damping and quadrupolar excitation (QE).[111−114] Figure 1 shows the linearity of the detected cyclotron frequency with the number of ions in the ICR trap under different conditions. Given a time scale of 2−3 s for QE signals in the presence of nitrogen gas at a peak pressure of 10^{-5} followed by a few second pump-down prior to detection, such a system would be impractical for a typical proteomic experiment with a capillary LC system, as one acquisition scan would be comparable or greater than the LC elution peak width.

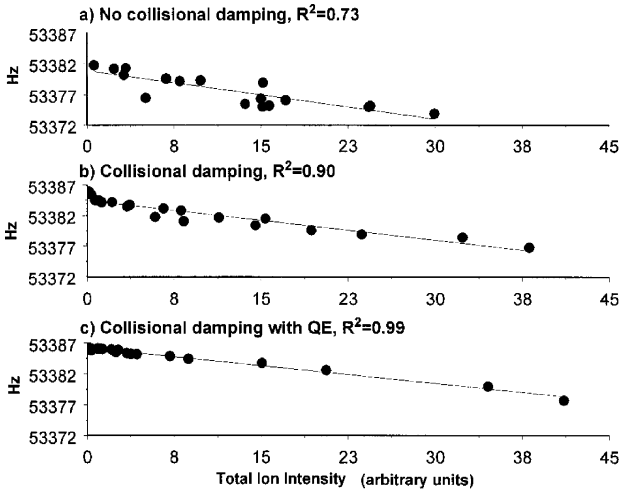


Figure 1. Observed frequency as a function of ion intensity for substance P measured over 109 laser shots on a 4.7-T FTICR instrument. (a) Ions captured with gated trapping ($R^2 = 0.73$). (b) Ions captured with gated trapping and collisional cooling with a pulsed buffer gas show improved linearity due to damping of the trapping motion ($R^2 = 0.9$). (c) Addition of quadrupolar excitation to the experimental sequence creates uniform pre-excitation conditions and provides the highest linearity in frequency versus intensity for MALDI-generated ions ($R^2 = 0.99$). (Reprinted with permission from ref 109. Copyright 1999 American Chemical Society.)

An alternative approach, deconvolution of Coulombic affected linearity (DeCAL),[115] was developed to account for the mass differences for different charge states of the same molecular species generated by ESI. Space-charge-induced frequency shifts were compensated by correcting the cyclotron frequencies to minimize the errors in the deconvoluted spectrum of the multiple charge states of a peptide. For positively charged ions, the molecular weight (*M*) and cyclotron frequency (*f*) were governed by the equation

$$M = \left(\frac{kB}{f_n - \Delta f}\right)n - n(M_c) \tag{10}$$

where *B* is the magnetic field; *k* is the proportionality constant; and *n* and M_c are the number of charges and the mass of the charge carrier, respectively. This procedure improved the average mass error of peptides that resulted from tryptic digestion of bovine serum albumin to 3.6 ppm from 113.9 ppm. Some of the limitations of this method pertain to the need for detecting multiple charge states of a peptide in the same spectrum, which may not be the case in a proteomic experiment, as well as to the assumption that the frequency shift (Δf) is constant over the *m*/*z* range.

All of the aforementioned corrections tend to account for the total space charge accumulated in the ICR trap. Frequency shifts caused by ion cloud interaction in the ICR trap (i.e., "local" effects) were proposed to be corrected for as follows:[106]

$$\frac{m}{z} = \frac{c_1}{(\omega_+ - \delta\omega_c)} + \frac{c_2}{(\omega_+ - \delta\omega_c)^2} \tag{11}$$

where ω_+ is the measured cyclotron frequency and c_1, c_2, and $\delta\omega_c$ are calibration constants, with the latter being dependent on the cyclotron radius. Importantly, at a fixed cyclotron radius, the mass calibration determined by eq 11 converges to that of eq 7. Only at varying cyclotron radii does the difference between the uniformly charged ellipsoid model[98] and the model of two interacting ion clouds[106] become significant.

In accord with earlier predictions,[106] lower- and higher-abundance species detected in the same spectrum were experimentally found to experience different frequency shifts, such that more intense peaks had positive frequency shifts, while less intense peaks revealed negative frequency shifts.[116] This observed phenomenon correlated with the concept that the space charge associated with an ion cloud consisting of particles of the same *m*/*z* cannot influence the center-of-mass motion of the cloud.[117] Invoking "local" frequency shifts resulted in a decrease in the mass measurement error by a factor of 3, though not fully compensating the systematic frequency shifts over the entire *m*/*z* range.

2.1.2. Internal Mass Calibration

Conventional internal calibration procedures imply that, when measured in the same spectrum, internal standards and analytes experience similar frequency shifts (only total space charge is considered) and the space-charge-induced term can be canceled out. Internal calibrants are introduced into an ESI-FTICR mass spectrometer as either (1) calibrants that coelute with analytes in a sample solution delivered to a

single ESI emitter or (2) calibrants and analytes that are spatially separated in a dual ESI source.[118,119] An internal calibrant-free calibration method with a single ESI source by using fragment ion information (e.g., fixed mass difference between two neighboring peptide fragment ions) has also been reported.[120] Although incorporation of analytes and internal standards simultaneously in the same solution has previously been accomplished successfully,[121,122] caution should be taken with respect to the hydrophobic properties of the internal standards to avoid analyte suppression in the ESI plume.

Internal calibration with a single ESI emitter initially was aimed at improving FTICR mass accuracy for the study of large biomolecules.[123,124] Internal calibration with MALDI represents a greater challenge for accurate mass measurements due to the broader (than ESI) *m*/*z* range and preferential ionization of lower charge states.[125,126] The distribution of errors for tryptic peptides digested from bovine serum albumin was studied using both nanoLC-microESI and MALDI sources.[127] Figure 2 shows the distribution of mass errors obtained using external and internal calibration modes for both MALDI and ESI experiments. The standard deviation for the distribution of errors in the nanoLC-microESI experiments was found to be ~1.2 ppm for both internal and external calibration, while the results from MALDI data

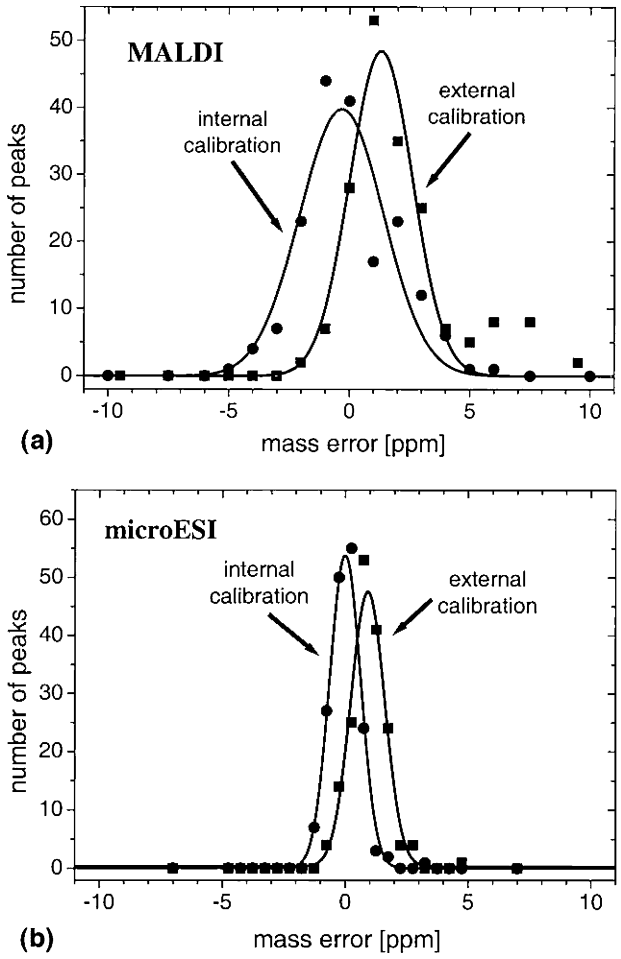


Figure 2. Distributions of mass errors with applied Gaussian functions for internally and externally calibrated data for (a) MALDI measurements and (b) NanoLC-microESI measurements on a 7-T FTICR instrument. Mass errors were calculated from all spectra obtained with (a) 1.5−50 fmol of analyte and (b) 1−50 fmol of analyte. (Reprinted with permission from ref 127. Copyright 2003 Elsevier.)

revealed standard deviations of ~3 ppm. Though internal calibration corrected the distribution means to 0 ppm, the broader error distribution observed in MALDI experiments could not be improved with internal standards.[127]

A dual ESI source coupled with FTICR has been demonstrated to internally calibrate precursor and fragment ions of oligonucleotides.[118,128] An improved ESI assembly allowed the ion population to be controlled by altering the hexapole accumulation time for the internal calibrants and analyte. The switching time between two emitters was <50 ms, and a mass accuracy of 1.08 ppm was achieved in direct infusion experiments with bradykinin.[129] An alternative method of introducing the sample from the dual ESI emitter source was reported in capillary LC-FTICR (3.5 T) experiments that employed automated gain control (AGC).[130] Both analyte and calibrant were concurrently infused into a dual-channel electrodynamic ion funnel[131] so that the calibrant injection time was independently controlled by gating an "ion disruptor" plate in the ion funnel. In conjunction with external calibration, the capillary LC-ESI-AGC-FTICR provided a ~10-fold increase in the number of tryptic peptides identified from a bovine serum albumin sample as compared to the number obtained with fixed ion accumulation and external calibration methods.[131] The standard deviation of the mass measurement errors for the internally calibrated tryptic peptides decreased on average by a factor of 2 compared to that of the same peptides identified with external calibration.

In contrast to a direct infusion experiment where ion populations can be controlled reasonably well, the number of ions generated in an LC-MS experiment varies drastically over the entire course of the LC separation. FTICR can provide extremely high mass precision and MMA, which is best for trapping nearly constant, relatively small, and well-controlled ion populations in an ICR trap.[97,125] However, the protein concentrations of interest in proteomics studies can vary by more than 10 000-fold[6] and produce an even larger variation in relative ion abundances at the peptide level, thus exceeding the FTICR dynamic range of measurement. As a result, the use of capillary LC for separating complex proteolytic digests in conjunction with FTICR poses a major challenge with regard to obtaining accurate mass measurements. A standard means of improving mass accuracy in analysis of a system with a broad dynamic range is to (1) increase the magnetic field for FTICR,[83,132,133] (2) data-dependently maintain ion populations in the ICR trap at levels lower than the threshold at which the nonlinear frequency shifts occur[131,134,135] (e.g., by employing AGC in the external trap), and (3) apply internal calibration.[131,136] Figure 3 shows the distribution of the error values for an ESI-FTICR (11.5 T) mass spectrum of a complex polypeptide mixture that resulted from tryptic digestion of bovine serum albumin. The improved cyclotron frequency stability and reduced frequency shifts at the higher magnetic field enabled signal averaging without degrading MMA.[137]

A variation of internal calibration based on a multidimensional recalibration approach that utilizes existing information on the likely composition of a mixture has been recently reported.[138] This method takes into account the variable conditions of mass measurements and corrects the mass calibration for sets of individual peaks binned, for example, by the total ion count for the mass spectrum, individual peak abundance, *m*/*z* value, and the LC separation time. The multidimensional recalibration approach statistically matches measured masses, to a significant number of putative known

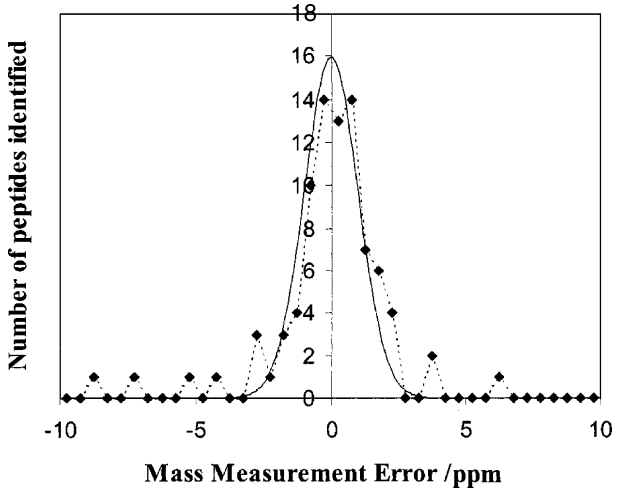


Figure 3. Distribution of error values observed between +10 and −10 ppm for the data of a complex polypeptide mixture resulting from tryptic digestion of bovine serum albumin. The dotted line represents the interpolation curve for the experimental error distribution, while the solid line shows the best fit for the experimental data with a Gaussian distribution. The large majority of error values fall near zero, and the distribution of error values closely resembles that of a normal error distribution with a standard deviation of 1 ppm. (Reprinted with permission from ref 137. Copyright 1999 American Chemical Society.)

species that are likely to be present in the mixture (i.e., having known accurate masses), to identify a subset of the detected species that then serve as effective calibrants. Figure 4 shows

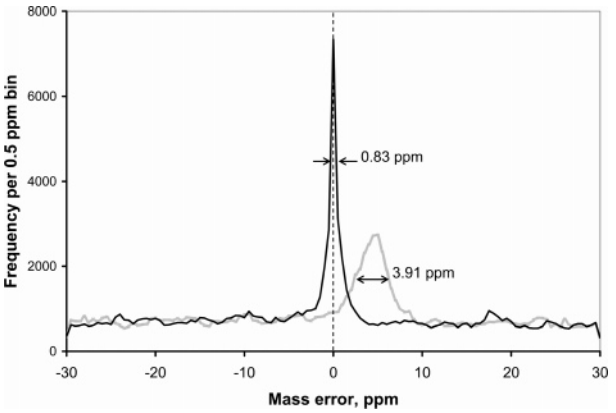


Figure 4. Mass accuracy histograms obtained for a *Neurospora crassa* fungus sample using an 11-T LC-FTICR MS. Results for instrument calibration (gray) and after recalibration (black). The number of calibration regions for TIC, *m*/*z*, and peak intensity is $10 \times 2 \times 10 = 200$. The systematic mass measurement error (i.e., histogram maximum position) is corrected from 5 to 0 ppm, and the mass error spread is improved from 3.9 to 0.8 ppm. The histogram maximum is increased >3 times, signifying a corresponding improvement in the certainty of identifications. (Reprinted with permission from ref 138. Copyright 2006 American Chemical Society.)

a mass accuracy histogram obtained using LC-ESI-FTICR (11 T) for analysis of a *Neurospora crossa* fungus sample. Note the systematic mass error is corrected from 5 to 0 ppm and the mass error spread is improved from 3.9 to 0.8 ppm. This recalibration can provide sub-ppm mass measurement accuracy for analysis of complex proteome tryptic digests and improved confidence in peptide identifications.[138]

Further improvements in FTICR mass accuracy could be achieved by combining the linearized excitation field[139,140] with the harmonic trapping field.[141] Trapping, excitation, and detection of carefully controlled ion populations in such a trap would be less dependent on space-charge and could potentially increase the mass accuracy of proteomic measurements with a capillary LC-FTICR instrument to routine sub-ppm levels.

2.2. Orbitrap Mass Spectrometry

The principles of ion trapping in electrostatic fields were described by Kingdon in 1923.[142] Orbital trapping was experimentally studied using an elaborate electrode shape (i.e., "ideal Kingdon trap") and ion spectroscopy.[143] The concept of ion trapping in a 3D electrostatic field was elegantly revised in a new type of analyzer that used ion axial oscillation and image current detection for high-performance mass analysis.[76,144] The electrostatic potential distribution within such a device is governed by[145]

$$U(r,z) = \frac{k}{2}\left(z^2 - \frac{r^2}{2}\right) + \frac{k}{2}(R_m)^2 \ln\left[\frac{r}{R_m}\right] + C \qquad (12)$$

where r and z are the cylindrical coordinates, k is the field curvature, C is a constant, and R_m is the characteristic radius.

Given polar coordinate (r,φ,z) treatment of eq 12, ion motion in the polar plane (r,φ) is decoupled from the ion trajectory along the z-axis. The latter parameter represents oscillatory motion with the characteristic frequency:

$$\omega_z = \sqrt{\frac{q}{m}k} \qquad (13)$$

where m/q is the mass-to-charge ratio of the ion. Ion motion in the polar plane (r,φ) is described by the radial oscillation and rotation frequencies:[76]

$$\omega_{rad} = \omega_z\sqrt{\left(\frac{R_m}{R}\right)^2 - 2} \qquad (14)$$

$$\omega_r = \omega_z\sqrt{\frac{\left(\frac{R_m}{R}\right)^2 - 1}{2}} \qquad (15)$$

Only the axial oscillation frequency, ω_z, is completely independent of the energy and position of the ions, thus promoting the "ideal Kingdon trap" to an orbitrap mass spectrometer.

Similar to FTICR, the orbitrap acquisition system is based on image current detection followed by fast Fourier transform. Since an ion cloud in the orbitrap tends to maintain coherence throughout the transient along the extended z-axis, the instrument has been claimed to have greater trapping volume and be less susceptible to the space-charge-induced frequency shifts than FTICR.[76] When the orbitrap was coupled to an ESI source, a mass resolving power of 150 000 (full width half maximum) and mass errors of <4 ppm were demonstrated in direct infusion experiments with a mixture of polymers and peptides.[146] When incorporated with a linear ion trap (LTQ)[147] and interfaced via a C-trap (an RF-only quadrupole shaped in the form of the letter "C" that accumulates and stores ions),[148] the orbitrap has been used in capillary LC experiments to characterize complex Lys-C digests of parotid saliva.[149] Orthogonal ion injection from the C-trap into the orbitrap constituted a significant advance over the axial injection method,[146] and mass resolving

power in excess of 40 000 and mass accuracies of <2 ppm were reported.[149] In another recent experiment, the mass accuracy of a hybrid LTQ/orbitrap (LTQ-Orbitrap, Thermo Electron) in both MS and tandem MS modes was evaluated. Background ions that originated from ambient air were first transferred to the C-trap, and then analyte ions along with internal standards were injected into the orbitrap for further analysis and internal calibration. Both precursor ions and fragments were identified with an average absolute deviation of 0.48 ppm and maximum deviations of <2 ppm.[150] Complete characterization of the orbitrap mass accuracy as a function of the number of trapped ions has recently been performed at the manufacturer's site (Thermo Electron, Bremen, Germany). Figure 5 shows the distribution of mass errors

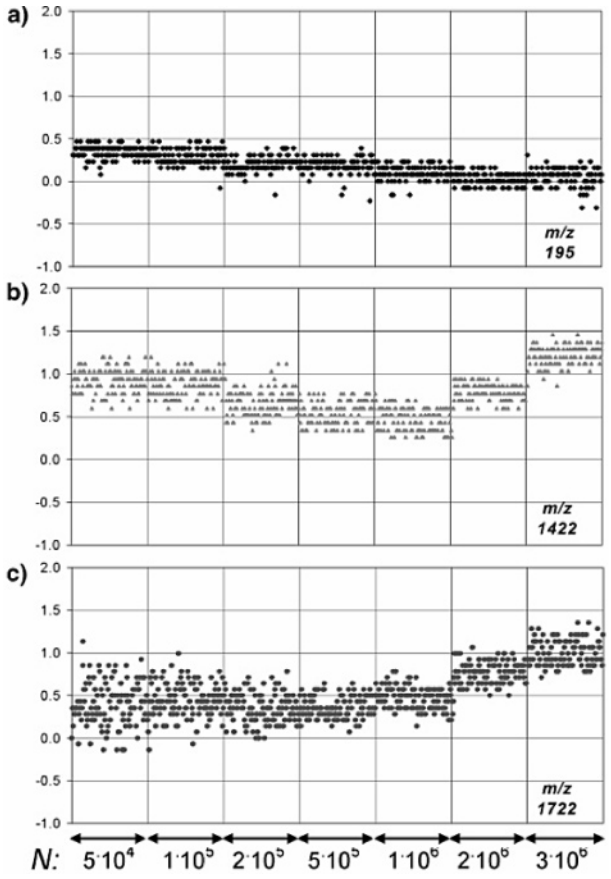


Figure 5. Mass errors plotted for different *m*/*z* values as a function of AGC target value *N* with the mass peak of the MRFA peptide (*m*/*z* = 524.2649) used as an internal calibrant at *R* = 30000: (a) *m*/*z* = 195.0876, (b) *m*/*z* = 1421.9778, and (c) *m*/*z* = 1721.9587. (Reprinted with permission from ref 152. Copyright 2006 American Chemical Society.)

for the analytes covering an *m*/*z* range of ~1500 as a function of the AGC target value. Given a dynamic range of ~5000, mass accuracies of better than 5 and 2 ppm were reported for external and internal calibrations, respectively.[151,152]

Since the release of the first commercial instrument (LTQ-Orbitrap) in 2005, orbitrap technology has increasingly been gaining ground in proteomics research. High mass accuracy (<2 ppm), high resolving power (>40 000), high sensitivity (<5 nM), increased dynamic range (~5000), impressive reliability, and low maintenance cost have made this instrument an attractive platform for a number of biological applications.

2.3. TOF Mass Spectrometry

The concept of separating ions with different *m*/*z* values via TOF was originally proposed in 1948,[153] but the first TOF mass spectrometers of any practical interest were not developed until the early 1950s.[154] The main characteristics of a TOF mass spectrometer include (1) unsurpassed analysis speed; (2) the ability to detect a complete mass spectrum in a single acquisition; (3) in principle, no upper limit for ion detection (in practice, limited only by detector efficiency for high mass ions), and (4) high sensitivity. In their seminal publication, Wiley and McLaren[154] described ion spatial and energy spreads as the main factors that affect TOF MS resolution, and they proposed time-lag energy focusing to narrow down an ion's energy distribution. These ideas continue to be exploited, as evidenced by the related development of delayed extraction (DE)[155−158] MALDI-TOF analysis. The effects of the initial energy (or velocity) distribution on TOF mass resolution were significantly reduced with the introduction of an ion mirror.[159] Using a two-stage ion mirror, second-order time focusing was achieved with a mass resolving power up to 35 000,[160] although this enhancement in mass resolving power could only be obtained in a narrow range of the mass spectrum. Development of new pulsed laser ion sources (e.g., MALDI) combined with reflectron TOF (RETOF) MS renewed the interest in TOF technology in a number of applications.[161]

Another important development involved orthogonal acceleration (oa) of ions into the TOF MS (oaTOF)[74,75] that enabled the coupling of a continuous ion source (e.g., ESI) to inherently pulsed TOF analyzers.[163−165] In oaTOF MS, an ion cloud is extracted to the TOF drift tube in a direction orthogonal to its initial trajectory so that only the ion velocity distribution in the plane perpendicular to the source axis contributes to the initial velocity spread. This initial velocity distribution along the TOF axis (and orthogonal to the source axis) translates to an ion cloud temporal spread (i.e., turnaround time[154]) in the reflectron object plane and cannot be compensated by the ion mirror. The ion turnaround time, Δt, is a major contributor to the overall peak width in oaTOF and is typically reduced to a few nanoseconds by increasing the extraction field and introducing efficient collisional damping prior to ion introduction into the oaTOF extractor.

$$\Delta t = \frac{2\sqrt{2mU_0}}{Eq} \tag{16}$$

where U_0 is the initial translational energy, m/q is the mass-to-charge ratio of the ion, and E is the electric field in the ion extractor. Given a proper instrument design, oaTOF MS is capable of achieving a mass resolving power of ~15 000−20 000 in a single pass,[166,167] and >50 000 in multiple-pass instrument.[168]

As TOF MS has been the sole subject of a monograph[169] and several review articles,[170,171] the reader is referred to these publications for additional information. The discussions below primarily focuse on obtaining accurate mass measurements with TOF MS as applied to proteomics.

2.3.1. MALDI-TOF

In idealized TOF MS, an ion with zero initial velocity has a time-of-flight proportional to the square root of its mass.[172,173]

$$\sqrt{\frac{m}{z}} = At + B \quad (17)$$

where A and B are the instrumental constants. By using eq 17 for internal calibration, mass measurement accuracies of 5−70 ppm with average values of 30−50 ppm were reported in "time-lag focusing" (i.e., delayed extraction) MALDI−linear TOF MS experiments with a mixture of peptides.[156] In the following study by the same group, a MMA of 80 ppm or better was demonstrated in DE-MALDI TOF analysis of poly(ethylene glycol) (repeat unit mass of 44) of mass up to 25 000 units and poly(styrene) (repeat unit mass of 104) of mass up to 55 000 units.[174] Systematic evaluation of the mass accuracy of DE MALDI was conducted with a PerSeptive Biosystems Voyager Elite XL, RETOF MS.[175] The utility of the calibration equation (eq 17) was verified under different experimental conditions, including flight time variations from a single spot (<8 ppm) and different sample spots (<10 ppm), different delay times between the laser pulse and the extraction pulse (<4 ppm), and different pulsed acceleration voltages (<8 ppm). Figure 6 shows the distribution of mass errors of several peptides obtained from a single sample spot (A) and six different sample spots (B). Note that, in TOF MS, time-of-flight variations can translate to 2-fold greater mass errors. Interpolation of the calibration function obtained with internal standards yielded more accurate results than extrapolation to a higher m/z range. This discrepancy was related to mass-dependent kinetic energies (i.e., the initial velocity distribution) and an energy deficit that arose from ion collisions with neutrals in the sample plume above the surface.[175] This finding was consistent with an earlier report that a broad distribution of initial velocities for ions produced by low-pressure MALDI imposes a major limitation on the achievable mass accuracy.[176]

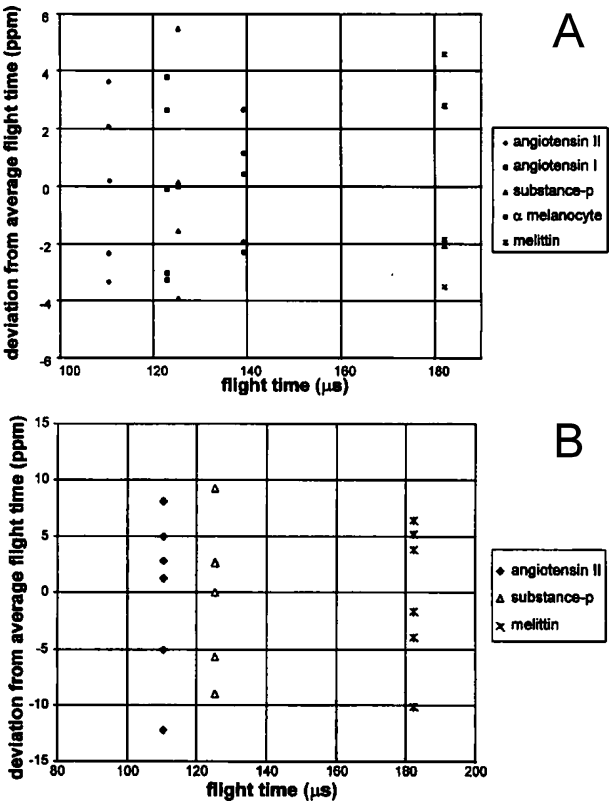


Figure 6. (A) A plot of the variation in the flight time (expressed as parts per million) of analyte ions taken from a single sample spot. (B) A plot of the variation in the flight time (expressed as ppm) of analyte ions taken from six different sample spots. (Adapted with permission from ref 175. Copyright 1996 Elsevier.)

To reduce the velocity distribution effect and couple the MALDI source to an oaTOF MS, orthogonal and on-axis injection of MALDI ions at an elevated pressure of 70 mTorr were developed.[177] Collisional cooling of MALDI ions in an RF quadrupole ion guide produced a parallel ion beam of small cross section and reduced the energy spread. Using eq 21 with substance P and melittin, a mass accuracy of 30 ppm or better was achieved for ions up to a mass of at least 6000 Da. The advantages of a higher-pressure MALDI source[178−180] included mass-independent calibration and the nearly complete decoupling of ion production from the mass measurement, features that affect the reproducibility and mass accuracy of a high-vacuum MALDI source. Higher-pressure MALDI-oaTOF was further evaluated in both MS and tandem MS modes, and a mass accuracy in the range of 10 ppm was obtained for both the precursor and fragment ions.[181] In addition to peak centroiding and long-term voltage fluctuations caused by temperature drift, accurate mass measurements of low-intensity signals in MS/MS experiments were observed to be determined by counting statistics, so that the statistical error would be $\sigma/\sqrt{N}$, where σ is the peak width and N is the number of ion counts. Given a mass resolving power of 10 000 for the parent ion, which corresponds to a peak width of 100 ppm [full width at half-maximum (fwhm)], 16 ion counts would result in a statistical error of ~10 ppm.[181] With improved implementation of the MALDI source interface to the TOF section via a collisional focusing ion guide, the instrument provided a uniform mass resolving power of 18 000 and a mass accuracy of 2 ppm in a single-point internal calibration of protein digest samples.[182]

2.3.2. ESI-TOF

The coupling of ESI to oaTOF MS[163−165] sparked a great deal of interest in TOF MS as a fast, accurate, high-resolution, and sensitive approach for peptide identification. CE/ESI high-accuracy TOF MS was employed to characterize small proteins, using peptide mapping. A reference solution containing L-methionyl-arginyl-phenylalanyl-alanine acetate (MRFA) and Ultramark 1621 was added to a mixture of proteins. Peaks for the ion electrophorogram of the tryptic peptide fragments and those for the two reference compounds (observed in mass spectra other than that of the fragment peaks) were averaged to obtain a single spectrum. In most cases, the error between the calculated and measured masses was <10 ppm. The measured masses of the protonated peptide fragments were then used to search against the EMBL database. The importance of mass accuracy in reducing the number of possible matches provided by the database appeared to be more pronounced when the number of peptides required for matching was only a few. For example, given four peptides selected for the match, an improvement from 15 to 10 ppm in mass accuracy resulted in a decrease in the number of matched proteins from 20 to 4.[183]

Although internal calibration has been shown to improve MMA,[122,184] the mixing of analyte with internal standards often results in analyte suppression, discrimination, and/or adduct formation. To avoid interactions between the analyte and reference standards, a dual-ESI sprayer coupled to a dual-nozzle in conjunction with oaTOF MS was designed and applied to obtain accurate mass measurements.[185] Observations indicated that the closer the bracketing reference peaks

were to the unknown, the lower the measured mass error. The standard deviation in mass error for seven samples in the mass range of 400−1000 Da was 3.3 ppm for the closet pair of bracketing peaks and 7.0 ppm for the second closest set of bracketing peaks. By comparing different methods for introducing calibrant into the mass spectrometer, the smallest errors resulted from the dual-ESI-sprayer dual nozzle (~3.5 ppm), followed by the methods in which the reference compound and sample were mixed either before or during the ionization process (~5 ppm). The highest errors were reported for sequential infusion of the internal calibrants and the analytes (~8 ppm).[185]

A number of other methods have been reported for obtaining accurate mass measurements. Two separate ESI sources for introducing the sample and a reference standard in the course of an LC separation were used for accurate detection of pharmaceutical compounds.[186] A method for diverting either the sample or the reference compound from the MS was reported as a multiplexed electrospray source.[187] With this method, a lock mass correction with leucine enkephalin yielded a mass accuracy of <3 ppm, and a mass accuracy of ~10 ppm could be achieved even at the edge of the detection limit [signal-to-noise ratio (S/N) of 6:1].[186] Another method reported using a dual-ESI-sprayer system to identify proteins by means of peptide mapping.[188] The tryptic digests of myoglobin (horse) were measured with a high-performance LC (HPLC)/dual-ESI-oaTOF MS instrument, with mass deviations that ranged from 0.01 to 7.67 ppm and ~75% mass deviations below 5 ppm.

The precision of a mass measurement with TOF MS depends mainly on ion statistics and is governed by[189]

$$\lambda_m = \frac{10^6}{CR\sqrt{N}} \tag{18}$$

where λ_m is the expression of the statistical error (e.g., the 95% confidence limit in ppm), C is an instrument constant, R is the resolving power, and N is the number of ions sampled in the measurement.

Precision is usually estimated indirectly, most often by incorrectly presuming that the error for all measurements is equal to the precision or the mean error for a set of reference masses. Based on eq 18, this is clearly not valid, as the precision of a particular mass measurement depends on the number of ions sampled and is likely to vary for every measurement. Mass measurement precision was established directly by making multiple measurements of the masses of interest and performing a statistical analysis of the data.[190] In an LC-TOF MS analysis of two pharmaceutical compounds using reserpine as a lock mass, the functional relationship between the precision and the number of sampled ions was well approximated by the linear fit ($R^2 = 0.9396$, intercept 0 ppm, slope 180.6 ppm). The error in the mass measurement was observed to increase significantly when the intensities for the analyte and the lock mass were significantly different. Increasing the signal rates improved the ion statistics and the precision, but resulted in a significant decrease in the accuracy of the mass measurement (up to 15 ppm).[190] A similar mass accuracy dependence on the intensity ratio of the target compound and the internal standard was recently reported for an LC-TOF MS analysis of $^{13}C_3$-caffeine[191] and a fully automated study with ~550 pharmaceutical compounds.[192]

Since an ESI interface to an oaTOF instrument provides efficient collisional focusing and minimizes the radial ion velocity distribution, calibration of eq 18 or its linear inversion is typically employed for correcting deviations of the experimentally measured masses from the calculated masses. To minimize the contribution of higher-order nonlinear effects, a mathematical procedure using multivariate fitting methods was developed[193] that involved a rigorous calibration model to eliminate the need for internal standards in a high-mass-measurement-accuracy LC-MS experiment. Two data processing methods were presented that corrected for systematic deviations: a peak fitting method using double Gaussian functions and a calibration method that takes into account the slight nonlinear response of the TOF analyzer. The model equation for the custom calibration technique is governed by

$$\left(\frac{m}{z}\right)_{cal} = a + b(\text{time}_{ret}) + c(\text{time}_{ret}^2) + d\left(\frac{m}{z}\right)_{ext} + e\left(\frac{m}{z}\right)_{ext}^2 \tag{19}$$

where $a-e$ are the fit parameters, m/z_{ext} is the externally calibrated m/z value, time_{ret} is the retention time, and m/z_{cal} is the calibrated m/z. The second and third terms in eq 19 account for a buffer change throughout LC separation and the associated space charge effects in the TOF MS extraction region due to more efficient buffer ionization at the end of LC separation. The systematic changes in MMA were observed over a time span of 1 h. A calibration solution that contained a mixture of several peptides was infused into the instrument before and after the LC separation, and three target peptides (e.g., from a tryptic digest of *D. radiodurans*) were also used to provide data points at intermediate retention times. The double Gaussian-multivariate method improved mass accuracy to 8 ppm (for serum albumin tryptic peptides) compared to 29 ppm, which was obtained using linear calibration and normal peak centroiding.[193]

Improvements in sensitivity (e.g., by using a microfabricated multiemitter ESI array) and mass resolving power (e.g., in a multipass reflectron) will further increase the mass measurement precision of oaTOF instrumentation. In addition, the use of analog-to-digital converter (ADC) based acquisition systems will reduce the dependence of the ion arrival time at the TOF detector on the ion abundance and further improve MMA in an analysis of a system with a broad dynamic range (e.g., the human proteome). Given the unsurpassed speed of analysis and the increasing need for high-throughput platforms for a number of clinical applications, TOF MS will continue to be an important asset at the forefront of proteomics research.

3. Accurate Mass Measurements in Proteomics

As a result of the continued technological advances in proteomics, various aspects of proteins, including structure, PTMs, relative abundance, localizations, and interactions with other molecules, can now be studied in unprecedented detail. We now focus the discussion on different proteomics approaches for peptide and protein identification, characterization, and quantitation and how accurate mass measurements enhance such analyses.

3.1. Peptide Mass Fingerprinting

Generally speaking, protein identification using a bottom-up approach is based on two processes: (1) generation of sequence information from proteins or peptide fragments

thereof and (2) inference of protein sequences (i.e., identification) by using such information. Until recently, protein sequencing was achieved by *de novo* sequencing using Edman chemistry, which is generally slow and low-throughput. Additionally, Edman-type sequencing typically requires the protein to be purified to near homogeneity, and a relatively large amount of protein is needed for a complete protein sequence. As a result of developments in MS technology and the increasing availability of annotated genomes for organisms, the information extracted from a protein or peptide by MS can now be correlated to a sequence database for identification in a fast, sensitive, and high-throughput fashion.

The most straightforward approach for identifying proteins using MS is PMF. Conceptually, the principle of PMF is quite simple: a group of peptides is produced from a protein by site-specific proteolysis (e.g., tryptic digestion) and their masses are accurately measured by MS to serve as unique mass fingerprints for the protein. The observed peptide mass fingerprints are subsequently compared to "virtual" fingerprints generated by *in silico* digestion of protein sequences stored in a database by a computer algorithm (e.g., MASCOT) that applies the same proteolytic specificity; the top-scoring protein is considered to be the identified protein. Among the drawbacks of this approach is that proteins that can be identified using PMF are limited to those whose sequences are at least largely known. The expressed sequence tag (EST)[194] databases are not suited for this purpose, because ESTs represent only a portion of a gene's coding sequence, which may not be long enough to cover sufficient numbers of observed peptides to allow unambiguous protein identification. Another obvious drawback with PMF is that it tends to result in ambiguous protein identifications for digests of unseparated protein mixtures in which different proteins give rise to peptides of similar mass. When working with such samples or with purified protein samples that originate from unknown species or species of which only limited genomic sequence information is available, sequence information is needed in addition to peptide mass information for unambiguous protein identification. Such information can be obtained by using MS/MS, which is discussed in section 3.2 below. The most common method of choice for protein identification with PMF has been the combination of 2-DE and MALDI-TOF, and many early proteomics projects relied on this method.[195−200]

The presence of unassigned masses in a typical PMF experiment detracts from the significance of probability based scores (e.g., the Mowse scores if using MASCOT), which may make the database searching outcomes indecisive. On one hand, data processing strategies that have been continuously developed to make better use of the information in PMF data sets and refine the peak list provide increased confidence in database searching.[54,201−206] Removing of the extraneous masses from the spectra, on the other hand, would allow enhanced database searching specificity to be achieved. Known contaminant masses (e.g., human keratin peptides and trypsin autolysis peptides) can be easily excluded from the PMF peak list using postprocessing tools. However, it is only recently that a strategy has been reported for removing the nonpeptide signals in the PMF peak lists based solely on the accurately determined monoisotopic masses,[207] since the monoisotopic mass of a peptide must fall within a predictable range of residual values.[53] As an example, a maximum error of approximately 15 ppm is required to avoid inappropriate rejection of a peptide with a nominal mass of 2001 Da, while sufficient resolution is also demanded to allow clear selection of the monoisotopic mass. Application of this strategy provided exponential improvements in the statistical significance and discrimination of PMF protein match results. Importantly, this scheme for removal of nonpeptide masses does not affect the post-translationally or artificially modified peptides.

In a sequence database, an increase in MMA results in decreased numbers of isobaric peptides for any given mass; this behavior is even more significant as the mass increases. As a result, peptide MMA is the most critical factor for protein identification using PMF.[35,52,208,209] At high MMA, a significant fraction of peptides that have the same nominal mass but different elemental and amino acid compositions can be removed, which will increase not only the speed but also the specificity of database searching. A TOF mass analyzer can potentially achieve a mass accuracy of 5 ppm by using internal calibration. However, the mass resolving power and mass accuracy of a linear MALDI-TOF instrument are constrained by a broad initial kinetic energy distribution[176] and mass-independent initial velocities[210,211] of MALDI-generated ions, which results in mass spectra with unresolved isotopic distributions. The use of higher-resolution and higher-mass-accuracy instrumentation (see section 2) significantly increases the confidence in peptide identification. Utilization of alternative data acquisition methods may further enhance protein identification in PMF. For instance, a simple procedure in which two sets of data are combined by using tuning conditions that favor low-mass ($m/z < 2000$) and high-mass ($m/z > 2000$) ions improves protein identification by 70% compared with the analysis of the same sample using a wide mass range acquisition on an HPLC-MALDI-FTICR instrument.[212] The importance of higher MMA is emphasized by the fact that although high accuracy significantly decreases the number of random matches to a database, some random matches can still be found, even at a mass accuracy of 6 mDa.[213] Moreover, peptides that differ in composition by one (or two or three) amino acids have a surprisingly high percentage of isomers: 10% (or 14% or 38%, respectively), excluding isomers that differ by leucine/isoleucine and assuming the 20 common amino acids have equal relative abundance.[69] Thus, it is still desirable to incorporate additional physical and/or chemical information (e.g., M_r, p*I*, hydrophobicity, proteolytic cleavage site) to achieve highly confident and unambiguous protein identification.

Adding other discriminating constraints can also increase the specificity of database searching. Site-specific chemical modification has been a common method for deducing the presence of specific amino acids in the peptide analyzed. For example, various mass shifts can be produced by alkylating a protein using different alkylation reagents. The cysteine content information is readily obtained by reacting sulfhydryl groups with a 1:1 mixture of unlabeled and stable isotope-labeled alkylation reagents, which can then be used to improve the protein identification process.[214] Moreover, mass defect labeling of cysteine by using a chlorine-incorporated alkylation reagent[215] or a novel reagent such as 2,4-dibromo-(2′-iodo) acetanilide[216] has been effective for improving identification by accurate mass measurement of labeled peptides. The natural isotopic distribution of chlorine or bromine encodes the cysteine-containing peptide with a distinctive isotopic pattern that allows for automatic screening of mass spectra (Figure 7).

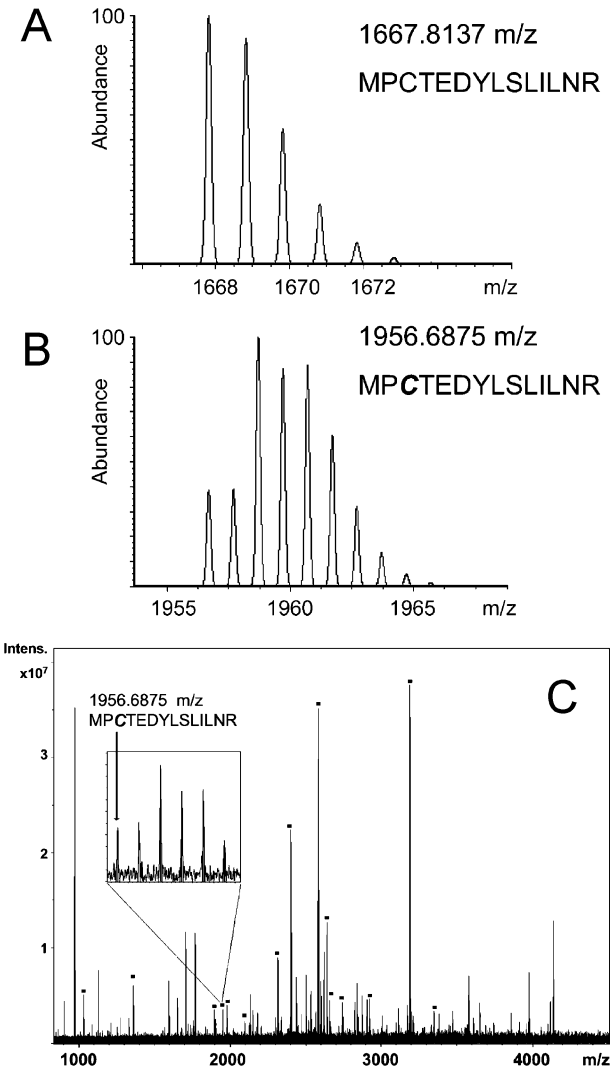


Figure 7. Calculated isotopic pattern for the peptide MPCTEDYLSLILNR from BSA (residues 445−458) (A) without and (B) with the dibromoacetanilide mass defect label. The MALDI-FTICR spectrum obtained of a BSA digest is shown in part C. Mass defect labeled peptides are denoted with a box. The inset shows a mass scale expansion of the peaks near *m/z* 1957, identified as the peptide MPCTEDYLSLILNR, whose predicted isotope pattern is shown in part B. (Adapted with permission from ref 216. Copyright 2006 American Chemical Society.)

3.2. LC-MS/MS Analysis of Peptide Mixtures

The use of MS/MS to generate sequence-specific spectra for peptides has been a popular approach for large-scale protein identification via automated sequence database searching.[46,49,217] The sequence-specific and information-rich fragment ion spectra generated in LC-MS/MS experiments can also be used for *de novo* peptide sequencing. Use of the partial amino acid sequence generated from Edman degradation has been a traditional approach to generate probes for isolating the gene coding for the protein from a gene library. Similarly, from an MS point of view, the amino acid sequence of even a relatively small peptide could potentially lead to unambiguous identification of a protein.

In the following sections, we discuss the importance of high MMA in large scale peptide identification, *de novo* peptide sequencing, and peptide PTM characterization using LC-MS/MS.

3.2.1. Increased Confidence in Peptide Identification

In general, a fragment ion spectrum is produced in three consecutive steps: (1) selection and isolation of the parent ion, (2) fragmentation of the parent ion, and (3) recording the fragment ion spectrum. Fragment ions can be generated through low-energy CID in a collision cell of tandem mass spectrometers [e.g., IT, TQ, quadrupole-TOF (Q-TOF)] and via fragmentation of a high-energy ion using post-source decay (PSD) in a MALDI mass spectrometer or using the magnetic sector or TOF/TOF mass spectrometers. The fragment ion spectra generated from high-energy ions typically contain ions that result from fragmentation of both the peptide side chain and the backbone. As a result, the spectra are complex and often difficult to interpret. In contrast, the low-energy CID spectra are dominated by N- and C-terminal fragments of peptide ions at the amide bonds, called *b* ions and *y* ions,[29,30] respectively. These spectra are of higher quality and more sequence specific.

Potentially, every peptide bond may generate a *b* or *y* ion. Therefore, an ideal peptide MS/MS spectrum would have two ladder-like ion series that start from the N-terminal and C-terminal of a peptide, respectively, and have identical ion intensities. While interpretation could be performed directly from an MS/MS spectrum, in practice, it is rare to see a perfect ladder of ions because, in addition to mass and charge, the optimal collision energy depends on the peptide sequence and tertiary structure and location of protonated sites.[218−220] As a result, not all peptide bonds have the same tendency to fragment under specific CID conditions; thus, while some fragment ions dominate fragmentation spectra, others are rarely seen. For example, the presence of a proline[221−223] or aspartic acid[224−226] residue in a peptide has been observed to frequently induce internal fragments that significantly alter the intensity of the fragment ions. Under typical LC-MS/MS conditions (ESI and low-energy CID), tryptic digests of proteins yield mostly doubly charged ions which undergo extensive and readily interpretable fragmentation; triply charged ions have doubly charged fragment ions intermixed with singly charged fragment ions, and singly charged ions typically do not undergo extensive fragmentation under low-energy collision excitation, confounding spectrum interpretation.[227] However, in a MALDI-IT instrument, singly charged ions can also be efficiently fragmented by devising an excitation scheme that enables the deposition of sufficiently large amount of energy.[228] All of these aspects must be taken into account for accurate interpretation of MS/MS spectra.[6,229]

A high-quality MS/MS spectrum contains rich (near complete *b* and *y* ion series) and constrained (ideally only *b* and *y* ions; no internal fragment ions or multiply charged fragment ions) sequence information regarding a peptide, which is often sufficient for unambiguous protein identification.[17,230] As a result, a complex protein mixture can be enzymatically digested and analyzed directly by LC-MS/MS without the need for prior purification of individual proteins. The strategy of using fully automated LC-MS/MS methods, such as "data-dependent" MS/MS,[231,232] which automatically selects ions for fragmentation based on the signals of a "preview" full-scan mass spectrum, in conjunction with algorithms (e.g., SEUQEST, MASCOT) that correlate the MS/MS spectra with sequences in a database has been commonly used in many proteomics studies. While modern tandem mass spectrometers, e.g., an LTQ, can generate a very large amount of MS/MS spectra for large-scale protein

identification, current database searching algorithms can only correctly assign peptide sequences to a small portion of the total MS/MS spectra. This shortcoming in assignments is mostly due to low spectral quality and/or to the selection of nonpeptide species for fragmentation, and to a lesser extent to the presence of modified amino acid residues and/or permutation in the sequence that cannot be predicted from the database. (Note, site-specific modifications, if present, can still be identified by devising an algorithm to anticipate such modification at specific residues.) Also, due to the complexity of the proteome, algorithms often fail to distinguish between true positive and false positive identifications (i.e., multiple good matches per MS/MS spectra).

Several recent developments address the need for accurate MS/MS peptide and protein identifications, including the use of statistical strategies that estimate FDR[58,60] and provide confidence matrices to the peptide assignments,[57] the use of multiple stage fragmentation,[66,67] application of other fragmentation techniques in addition to CID,[68] and new algorithms that utilize addition information from MS/MS spectra, e.g., fragment ion intensities.[233,234] The concept of "peptide sequence tags" has also been introduced to increase the specificity of peptide identification from MS/MS data.[33] A partial sequence can often be inferred from a short and easily identifiable fragment ion series. This partial sequence, together with the mass information of the fragments to the left and right side of it, constitutes a peptide sequence tag that is a highly specific identifier of the peptide and can therefore be used to search the sequence database for peptide identification.

High resolution and MMA have been leveraged successfully to improve peptide sequence identifications from MS/MS spectra.[149,150] Accurately measured mass information can significantly increase the specificity of database searching in two ways:[35,209] (1) on one hand, high MMA for precursor ions that is typically obtained in an MS prescan effectively reduces the number of possible candidate parent ion sequences that need to be matched against the fragment ion spectra; (2) on the other hand, high MMA in the MS/MS spectra reduces random matches of the fragment masses and hence decreases false positives. However, conventional tandem mass spectrometers, exemplified by the widely used 3-D IT (LCQ) and linear IT (LTQ) instruments, have limited mass resolution and MMA for parent ions in typical large-scale proteome analysis. Both parent ions and fragment ions can be measured at high MMA on several new MS/MS instruments, e.g., the Q-TOF,[235] hybrid LTQ/FTICR (LTQ-FT),[134] and LTQ-Orbitrap,[76,149,150] which are being increasingly used in proteomics applications. Note that, in a large number of experiments where these instruments have been used, an accurate mass of the precursor ion is collected but MS/MS is still carried out in regular resolution mode for speed and sensitivity reasons. Alternatively, a multiplexed MS/MS approach (i.e., dissociation of several species simultaneously in a single experiment) using FTICR has been advised.[236,237] The high MMA and resolution obtained in such analyses allow the fragments that arise from several parent ions to be assigned.

New MS/MS instruments have attracted tremendous attention in proteomics studies and also raised considerable concerns with regard to data quality and false positive identifications.[238−240] For example, in one of the early large-scale proteomics applications of Q-TOF, a MMA of better than 20 ppm for both the precursor and fragment ions was obtained for analysis of selected stages of the human malaria parasite *Plasmodium falciparum*.[241] To increase the peptide sequencing speed, the Q-TOF instrument can be operated in such a mode that all ions that enter the ion source are simultaneously fragmented *in situ*, and both precursor and product ions are measured at high mass accuracy in the TOF mass analyzer using as few as two scans.[242] This approach significantly improves the duty-cycle inefficiency that is inherent in a typical "data-dependent" MS/MS analysis; however, the MMA achievable is limited to 5 ppm for the precursor ions and 10 ppm for the product ions, even with internal calibrants. Multiplexed peptide fragmentation has been carried out using an FTICR mass spectrometer and provides both increased MMA and sensitivity.[236] When coupled to an on-line separation,[237] the utility of this approach has been demonstrated for high-throughput identification of tryptic peptides from large databases.[243] A so-called "patchwork peptide sequencing" approach that extracts sequence information from accurate masses recorded in the low-mass region of MS/MS spectra (m/z 60−400) also appears to be efficient for protein identification using the Q-TOF data.[244] In another study, monitoring of a_1 (that resulted from the neutral loss of carbon monoxide from the b_1 ion) or a_1-related ions in the low-mass region of Q-TOF MS/MS spectra of peptides labeled by 2MEGA (dimethylation after guanidination) provides an additional constraint for database searching and reduces false positive peptide identifications.[245] Additionally, the unique LTQ-FT and LTQ-Orbitrap hybrid instruments both have an LTQ and a highly accurate mass analyzer (FTICR and orbitrap, respectively) that can be operated either independently or concordantly to achieve high MS/MS speed and high MMA.[150,246]

Typically, MS/MS operation targets one specific ion species for CID. As a result, the fragment ion spectrum will exclude internal mass reference ions, which precludes achieving high MMA for unambiguous fragment ion mass assignments. However, in the LTQ-Orbitrap, ions accumulated in the LTQ can be transferred into a C-trap and collisionally damped there before being injected into the orbitrap for highly accurate mass measurement.[150] This feature has enabled a novel way of introducing the "lock mass" for real time calibration to compensate for drift in the electric field over time. A predefined number of the protonated electrospray ion of polycyclodimethylsiloxane (PCM-6) that is being generated during the electrospray process is accumulated in the LTQ and transferred to the C-trap. These ions can then be added to any spectrum for highly accurate mass measurement. The remaining mass error can be further improved by averaging mass measurements over the LC peak weighted by signal intensity. Better than 1 ppm MMA can be achieved using this approach.[150] Also, the MS/MS spectra obtained in the LTQ and orbitrap are comparable in terms of fragment ion pattern and intensity, but the MS/MS spectra recorded in the orbitrap (1 ppm MMA with the lock mass strategy) contain fewer noise ions than spectra recorded in the LTQ, presumably due to the high resolution of the orbitrap and its image current detection. When searching orbitrap MS/MS data with common database searching algorithms (e.g., MASCOT), the "delta score" that distinguishes the top hit from the next best matching peptide sequence has been noted to increase dramatically and at other times there is no second hit at all,[150] which indicates that the high MMA in the MS/MS spectra had a significant impact on the specificity of peptide identification. Currently,

most algorithms cannot make use of the extremely high MMA in MS/MS spectra for database searching, and thus, the scores for fragment matches do not increase. Developments in bioinformatics are starting to address this need; for example, the PeptideProphet algorithm[57] can now incorporate an accurate precursor mass as additional evidence for assigning confidence to a peptide identification.

The LTQ-FT and the LTQ-Orbitrap instruments have made routine accurate mass measurements for shotgun proteomics practical, and new strategies are emerging. For example, using the "selected ion monitoring (SIM)" scan[240] function over a narrow *m*/*z* range has been demonstrated to improve the MMA in LTQ-FT analyses. The use of AGC with this method enables even a small number of ions to be measured at high resolution ($R = 50\,000$) and high MMA (<1 ppm); however, the overall duty cycle is low. To investigate the benefits and costs of using accurate mass measurements in typical shotgun proteomics studies, a highly complex peptide mixture derived from yeast was analyzed on the LTQ-FT instrument. The FTICR part of the hybrid mass spectrometer was either not exploited, was used only for survey MS scans, or was used for acquiring SIM scans, and the numbers of confident peptide identifications were compared.[70] MS/MS analysis with high MMA was noted to provide slightly more peptide identifications (~10%) than analysis with more typical MMA; however, the excessive pursuit of extremely high MMA can be at the expense of the MS/MS acquisition rate, which could substantially decrease the sensitivity and sequence coverage in a typical data-dependent LC-MS/MS analysis. Further investigation showed that the benefits of high MMA were greatest for assigning spectra with low S/N values (i.e., low-abundance peptides) and for assigning generally low-quality phosphopeptide spectra (e.g., incomplete ion series, significantly altered ion intensity because of neutral loss), in which peptide identifications can be doubled.[70] Another benefit of using the high-MMA data was that the database searching time could be reduced by applying a narrower peptide mass search tolerance. In summary, combining the high MMA of precursor ions with fragment ion information obtained on LTQ-FT leads to high confidence (typically <1% FDR) in peptide and protein identification, as reported recently in a number of studies.[240,246−249]

3.2.2. De Novo Peptide Sequencing

A MS/MS spectrum automatically searched against a sequence database does not always lead to confident peptide identification, even for spectra with very good S/N and many fragment ions. The presence of splicing variants, protein isoforms, fusion proteins, or novel PTMs can all result in poor quality matches. Additionally, studying species with yet uncharacterized genomes is not possible with current database searching algorithms. An alternative strategy for interpreting such data either with or without minimal assistance from genomic data is *de novo* peptide sequencing. Historically, *de novo* sequencing by MS was performed via Edman degradation without using MS/MS by generating peptide ladders that differed in length by one amino acid and by measuring their masses using a MALDI instrument to "read" the sequence of the peptide based on the mass differences.[250] However, the low throughput and low sensitivity of this type of method limited its broad application in proteomics. MS/MS based methods are particularly attractive because a typical LC-MS/MS analysis can now generate tens of thousands of high-quality MS/MS spectra that could possibly be *de novo* sequenced provided enough informative *b* ion and *y* ion series peaks are present. Since both *b* ions and *y* ions may be present in a typical MS/MS spectrum, a key issue in *de novo* sequencing is that the sequence cannot be easily interpreted unless the directionality of the ion series can be determined. Various approaches, including chemical derivatization[28,251−254] and isotopic labeling,[255−260] have been developed to address this issue. For example, using a high-resolution Q-TOF instrument, the sequence of a peptide labeled with $^{16}O/^{18}O$ can be readily discerned due to the high quality of the MS/MS spectra.[255] There has also been active development of new bioinformatics tools for *de novo* sequencing of high-throughput proteome-wide LC-MS/MS data.[261−268]

De novo interpretation of MS/MS spectra derived from a typical highly complex tryptic digest proteomics sample is desirable. For marginal quality spectra (i.e., noisy data with only a handful of fragment ions present), the MMA of both the precursor ion and fragment ions is critical to the confidence of peptide sequence assignment; when the ion series are not complete, the interpretation draws heavily on the internal fragment ions in the spectrum, which is generally performed in manual spectrum confirmation and not included in a standard database search. Although the 20 common amino acid residues have distinctive elemental composition and masses (except for Leu/Ile), the combination of amino acid residues can yield the same mass number or even the same elemental composition (e.g., Gly + Gly = Asn; Gly + Ala = Gln). In order to distinguish different combinations of amino acids, different degrees of MMA may be required. A low-MMA instrument with unit resolution may not be able to discriminate a 1 Da difference, e.g., Asp vs Asn or Glu vs Gln. A moderate MMA of <30 ppm is required to distinguish between the sequences "Thr-Thr-Tyr" and "Asp-His-Leu" ($\Delta m = 11$ mDa), which is typically achievable by advanced TOF instruments. Thus, the presence of "gaps" in the ion series hampers manual attempts at spectra interpretation because the number of possible di-, tri-, and tetrapeptide combinations that "fit" the same gap could be enormous if the MMA does not provide sufficient specificity. At a MMA of 10 ppm, the *de novo* interpretation of MS/MS spectra from peptides with parent mass <1300 Da is practical using a hybrid strategy that employs *de novo* MS/MS interpretation followed by text based sequence similarity searching of a virtual database (i.e., matching the sequences deduced *de novo* to the sequences in the database) rather than the entire genome database.[35] This virtual database can be generated on-the-fly to include only the set of amino acid combinations and all permutations of each combination that are dictated by the accurately measured masses of the parent ion and immonium ions.[35]

It is possible to interpret larger sequences by using either a more sophisticated approach to reduce the number of sequence permutations that need to be examined or significantly improved MMA (e.g., <1 ppm). In particular, extremely high MMA is now achievable on much more user-friendly FTICR and orbitrap instruments;[70,150,240] however, bioinformatics tools that take full advantage of exact mass information for high confidence are still far from mature. To address this need, a new strategy for non-database-assisted peptide sequencing has recently been reported,[269] which involves a critical first step to determine the amino acid composition based on the accurately measured masses (peptide composition analysis) obtained on an LTQ-FT

instrument. In the second step, termed composition based sequencing to be distinguished from conventional *de novo* sequencing, the amino acid sequences of the peptide are determined by scoring the agreement between expected and observed fragment ion signals of the permuted sequences. In this strategy, the efficiency of permutation and calculation of all possible amino acid sequences, which is the key to the overall success for high-confidence peptide sequencing, depends strongly on the MMA achievable in the analysis.[269] A similar approach that utilizes a peptide composition look-up table indexed by residual mass and number of amino acids has been applied for *de novo* sequencing of peptides using MALDI-TOF/TOF data.[270] Obviously, limitations of these strategies are still present for large peptides, and their application in large-scale proteomics studies needs to be demonstrated.

Unlike the database searching algorithms used in typical MS/MS experiments, *de novo* sequencing algorithms deduce and score all possible peptide sequences using only sequence resources available in the spectra. Therefore, the *de novo* interpretation of MS/MS spectra often generates ambiguous or partial sequences due to insufficient fragment ion information or a too complex fragment pattern and due to the inability to distinguish certain amino acid residues at a specific MMA. However, these sequences can be used to drive complementary sequence homology searches[265,271] (e.g., FASTA and BLAST), providing independent interpretation of the MS/MS spectra that could validate the candidate sequences that rely on matching fragment ion patterns.[35,272] For example, a recent study using this strategy was able to rapidly assign (confirm or reject) more than 70% of peptide identifications of borderline statistical confidence from a MASCOT search, without manual inspection of the raw spectra.[272] However, the performance of this approach is inherently limited by the availability of meaningful candidate sequences produced by the *de novo* sequencing algorithms.

The success rate of *de novo* sequencing can be further improved by using the complementary fragmentation techniques (CID and ECD)[266,273,274] or by performing two consecutive stages of MS fragmentation,[66] since the additional and complementary fragment ion information provided by these techniques can be combined or correlated (e.g., MS/MS and MS^3)[275] for more exclusive peptide identification. Particularly, utilization of the new generation hybrid MS instruments (e.g., LTQ-FT) for *de novo* sequencing is desired, because they can provide not only more complete fragment ion information for significantly improved peptide identification but also the high mass accuracy (e.g., better than ± 0.04 Da)[266] necessary for obtaining low FDR in proteomics-grade *de novo* sequencing. In a comparison of peptide *de novo* sequencing using high-MMA data and low-MMA data, it has been shown that the percentage of error-free peptide identifications increases from approximately 30% for traditional MS instruments (e.g., LTQ) to 90% for precision MS instruments (e.g., LTQ-FT).[276]

3.2.3. Characterization of Post-translational Modifications

A proteome is not the product of the direct translation of gene sequences into protein sequences. Instead, many proteins have been post-translationally modified (some heavily) to be able to function properly and/or to play a role in cellular events; for example, reversible protein phosphorylation is a key regulatory mechanism in signal transduction.[277] Thus, characterization of PTMs is of great importance for developing an understanding of biological processes. However, analysis of PTMs poses significant challenges compared with conventional techniques for a number of reasons, which include the modification rate is high (e.g., the extent of modifications in the human proteome has been estimated to be one PTM per amino acid on average[278]), PTMs are often present at low stoichiometry, PTMs are frequently labile, different types of PTMs or multiple PTM sites may reside in the same region of the protein, and some PTMs have less defined structure (e.g., O-glycosylation).

Due to its high sensitivity, high accuracy, and versatility, MS has been used as a primary tool in the proteomics quest to study PTM and cellular regulatory mechanisms.[279,280] MS/MS analysis is particularly useful for this task because it can simultaneously identify not only the type of PTM present but also the accurate PTM site(s). Recently, the accurate and large-scale identification using MS of a number of important PTMs, such as ubiquitination[281] and sumoylation,[247,282,283] to name a few, has been reported. Developments in bioinformatics now allow the search of all types of PTMs at once without even knowing which PTMs exist in nature by using spectral alignment[284−286] or *de novo* interpretation.[268,287] The database searching speed, which typically increases linearly with the increase in database size and exponentially with the number of PTMs simultaneously considered using traditional database search algorithms (e.g., SEQUEST), can also be significantly improved by using the concept of spectral alignment.[285,288] In addition, the use of a spectral network constructed by aligning spectra from overlapping peptides can allow analysis of all correlated spectra at once, thus increasing the confidence of peptide and PTM identifications.[288] Furthermore, complementary fragmentation techniques (CID/ECD) using "precision mass spectrometry" have been suggested for high-confidence identification of unmodified and modified peptides,[276] and bioinformatics tools for using the accurate mass data and the combined CID/ECD datasets are becoming available.[274,289,290] Due to the limited scope of this review, we use phosphorylation and glycosylation as examples below to illustrate how accurate mass measurements can aid in detecting and identifying PTMs in large-scale proteomics applications, as the principles used in such analyses can be similarly applied for the characterization of other PTMs.

Protein phosphorylation/dephosphorylation catalyzed by various protein kinases/phosphatases often serves as an on/off "switch" in many important cellular events. While it is well-known that the most common phosphorylation sites are serine, threonine, and tyrosine residues, typically only less than 1% of the total identified peptides from a proteome analysis appear to be phosphorylated and post-translationally modified, and tyrosine phosphorylation only represents 0.05% of the total phosphorylation events in the cell.[291] The highly transient and dynamic nature of this PTM makes its proteome-wide characterization extremely challenging. Since protein is phosphorylated by forming phosphate ester bonds with hydroxyl side chains of Ser, Thr, and Tyr residues, a mass shift of $+80$ Da accounts for one phosphorylation site. Thus, the most straightforward method to identify phosphopeptides in a mixture of predominantly nonphosphopeptides is to track the peptide mass pattern before and after phosphatase treatment, which can be easily carried out on a MALDI-TOF instrument where the phosphatase reaction can take place *in situ* on the sample plate.[292,293] However, the efficiency of this method decreases as the sample complexity

increases. Moreover, using just the mass information for identification of phosphorylation may not be conclusive unless a substantially high MMA can be achieved in the analysis to distinguish this PTM from the others. For example, high mass accuracy is required to distinguish phosphorylation from sulfation ($\Delta m = 9.5$ mDa).[294,295] High-MMA measurements are also particularly attractive for phosphopeptide identification, because phosphorus has a distinctively large mass defect relative to H, C, and O (~0.03 Da). At 0.1 ppm MMA, >80% of the 2 kDa yeast phosphopeptides can be identified solely on the basis of their masses.[51] A number of CID based approaches have also been exploited to detect phosphopeptides. In-source CID uses a high orifice potential during the negative ion mode scan of the low-*m*/*z* range and then a reduced voltage that does not cause fragmentation while the high-*m*/*z* range is scanned. Phosphate-specific ions (e.g., 79 Da for PO_3^-) that are generated at high-voltage conditions can be monitored for the detection of phosphopeptides.[296] Similarly, methods have been developed for neutral loss scanning (monitoring the neutral loss of H_3PO_4, 98 Da)[296,297] and precursor ion scanning (monitoring the loss of PO_3^-)[298−300] using TQ instruments. These methods are all very useful for detecting the presence of phosphopeptides in a complex and unseparated peptide mixture; however, the lack of sequence information often limits their ability for unambiguous assignments of the phosphorylation sites.

To achieve large-scale and high-throughput protein phosphorylation analysis, pre-enrichment of phosphopeptides through immobilized metal ion chromatography (IMAC)[44] or strong cation exchange (SCX) chromatography[67] is typically coupled with automated LC-MS/MS analysis using low-energy CID in an IT instrument. Phosphopeptides are subsequently identified by specifying dynamic modification of +79.9663 Da on Ser, Thr, and Tyr residues during the automated database search. As mentioned earlier, filtering criteria are needed to remove false positives and reach a desired precision. Typically, a match quality score (e.g., Xcorr in a SEQUEST search) and a score that distinguishes the top hits (e.g., ΔCn in a SEQUEST search) are used. Such criteria, although proven effective in global proteomics analyses,[58−60] generally lead to significantly reduced sensitivity in the phosphoproteomics experiments. Xcorr scores for phosphopeptides are often suppressed and score similarities are prevalent (thus, the ΔCn value is generally small),[292] probably due to the insufficient fragmentation and generally reduced fragment ion intensity compounded by the prevailing neutral loss phenomenon[292] and the complexity of the fragmentation pattern if multiple phosphorylation sites are present. Chemical derivatization of the phosphopeptides through β-elimination and Michael addition reactions can effectively alleviate neutral loss and often provides a means for either enrichment or quantitation.[301−305] However, the sample loss in this multiple-step reaction results in generally decreased sensitivity of detection, a critical element for phosphoproteomics analyses.

Not surprisingly, most of the reported phosphoproteomics studies were performed without chemical derivatization, which usually requires manual confirmation of the MS/MS spectra to ensure confident phosphopeptide identification. Fortunately, this bottleneck in phosphoproteomics analysis is about to be broken as a result of the recent development and assessment of alternative filtering criteria that include mass accuracy and tryptic state constraints and are capable of producing a substantial increase in precision without compromising sensitivity.[306] Specifically, these alternative filtering criteria are enabled by using proper search space selection in combination with high-MMA data (e.g., from LTQ-FT analysis). The use of a relatively broader search space (50 ppm), a postsearch strict mass deviation cutoff (within an 8 ppm window), and a fully tryptic requirement made it possible to distinguish correct from incorrect peptide spectral matches with only modest Xcorr filters and no ΔCn filters, which rescued many correct matches from the low Xcorr area while maintaining a low error rate (Figure 8). In

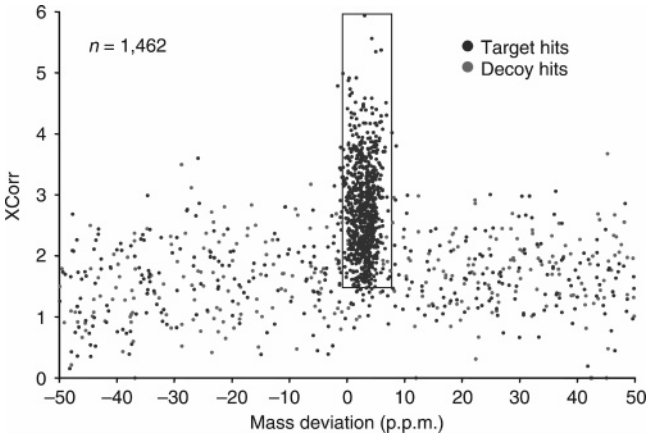


Figure 8. Effects of mass deviation as a filter for removing false-positive identifications. Correct tryptic phosphopeptide identifications distribute within an 8 ppm window and an Xcorr > 1.4 (boxed). False-positive identifications distribute evenly throughout the entire 50 ppm window. (Adapted with permission from ref 306. Copyright 2006 Macmillan Publishers Ltd.)

other protein phosphorylation analyses, accurate mass-driven analysis and rapid parallel MS/MS acquisition, a unique and practical strategy of commercial LTQ-FT and LTQ-Orbitrap instruments[70,150] that is independent of the signature neutral loss from phosphorylated amino acid residues, is very useful for unambiguously assigning phosphorylation sites and discovering new sites.[119,307] Moreover, the ETD technique, a combination of gas-phase ion/ion chemistry and MS/MS that induces fragmentation of the peptide backbone while preserving the labile PTMs (e.g., phosphorylation), has recently been made available.[308] ETD in combination with the strategy of parallel high-MMA MS and MS/MS acquisition is expected to provide unparalleled data quality for phosphoproteomics.

Protein glycosylation is another most common PTM; as high as 50% of proteins are estimated to be either lightly or heavily glycosylated.[309] More importantly, glycosylation plays a major role in cell−cell recognition, as well as in signaling through a reversible mechanism.[310] While the importance of protein glycosylation analysis has been well recognized, the progress made in this area has been slow, even with the tremendous advances in MS. Compared to phosphorylation analysis, complete glycosylation analysis requires not only identification of glycosylated proteins and peptides and glycosylation sites, but also illumination of the glycan structure. The latter requirement adds significant complexity to the analysis,[311] which is another topic entirely and is therefore not discussed further here. The presence of a glycan moiety in peptides can be selectively monitored using either precursor ion scanning in a TQ instrument or skimmer fragmentation in a single quadrupole instrument. The characteristic fragment ions that have been used com-

monly for detecting glycosylated peptides are the "reporter" oxonium ions of hexose at *m*/*z* 163.060, of *N*-acetylhexosamine at *m*/*z* 204.084, and of hexoylhexosamine at *m*/*z* 366.139.[312] However, other nonglycosylated peptides reportedly can also be detected using precursor ion scanning on a low-resolution TQ instrument, thus decreasing the specificity of such an analysis[313] because they produce other peptide-derived fragment ions (e.g., *a*, *b*, and *y* ions) that have the same nominal mass as the characteristic reporter oxonium ions. In the low-*m*/*z* range used by precursor ion scanning for selective monitoring of the reporter ions, there are a large number of amino acid compositions that have the same nominal mass. Some of these amino acids have very small differences in mass compared to the reporter ions; for example, the difference between the *N*-acetylhexosamine reporter ion and the a_2 ion of peptides containing an N-terminal sequence of "Ala-Cys" is only 3 mDa. Use of a high-resolution Q-TOF instrument has been demonstrated to be efficient for highly specific detection of glycosylated peptides through selective monitoring of the characteristic reporter ions, with only minimal interference from the peptide fragment ions.[314]

Typically, glycans can be attached to the peptide backbone via Ser or Thr residues (O-glycosylation) or via an Asn residue (N-glycosylation). To identify O-glycosylated peptides and O-glycosylation sites, a tag is typically introduced via β-elimination and Michael addition reactions, and the derivatized peptides can be readily identified by database searching.[315] The limited sample recovery and potential cross-reaction with phosphorylated peptides are the main disadvantages of this type of method. In contrast, the glycan on N-glycosylated peptides can be readily removed from the peptide backbone by incubating the peptide mixture with peptide N-glycosidase F (PNGase F), which converts Asn to an Asp residue while removing the glycan, which results in a mass shift of +1.008 Da of the formerly N-glycosylated peptide. Therefore, the use of hydrazide chemistry to enrich glycoproteins and the use of PNGase F to selectively release N-glycosylated peptides for LC-MS/MS analysis have been common for large-scale N-glycoproteome profiling.[45,316,317] The formerly N-glycosylated peptides are identified by database searching, using a dynamic modification of +1.008 Da for Asn residues. However, a typical tandem mass spectrometer used in such analysis (e.g., IT instrument) has only limited resolution, and the resulting data after database searching is often ambiguous (e.g., ΔCn score similarity in a SEQUEST search) for assigning N-glycosylation sites. This obstacle can be addressed by either accurately measuring the mass of the precursor ion or analyzing the same sample separately on a FTICR instrument to determine the number of N-glycosylation sites in the peptide.[317] To further differentiate between spontaneous deamidation and enzymatic deglycosylation as the cause of Asn to Asp conversion, it is necessary to apply an enzymatic deglycosylation reaction in 1:1 (v/v) $H_2{}^{18}O/H_2{}^{16}O$, from which a 2-Da mass increment can be introduced at the site of N-glycan attachment upon deglycosylation.[318] This increment can be easily monitored on a MALDI-TOF instrument; however, a high-resolution ESI-MS instrument is needed to detect the small difference in mass once the peptides are doubly or triply charged. The use of a high-mass-resolution MS instrument is also typically needed for characterizing highly complex glycoprotein digests.[319] Even with reversed-phase LC separation, the complex digest mixture may still contain overlapping isotope clusters of different molecular weight components, and high resolution may be essential for correctly identifying these species. In addition, while extracting the residual mass distribution of natural peptides from a protein database, the mass signals near the low-mass edge of the residual mass distribution have been observed to correlate with a high probability that the peptide is either a glycopeptide or contains one cysteine site, several cysteine sites, or a high number of Asp and/or Glu residues.[320] Glycosylation prominently lowers the residual mass value of a peptide, especially a small peptide, as a result of the high abundance of oxygen (15.995). Thus, the accurate mass and residual mass distribution can serve as unique indicators for glycopeptide identification and validation.

3.3. LC-MS Analysis of Peptide Mixtures

The highly accurate mass measurement capability using high-resolution MS has enabled broad applications of PMF approaches for protein/peptide identifications; however, these applications have often been limited to relatively simple peptide/protein mixtures. For example, with MMA of ~1 ppm, 85% of the peptides predicted from *S. cerevisiae* and *C. elegans* were expected to function as accurate mass tags (Figure 9).[51] This level of MMA could allow for "unique"

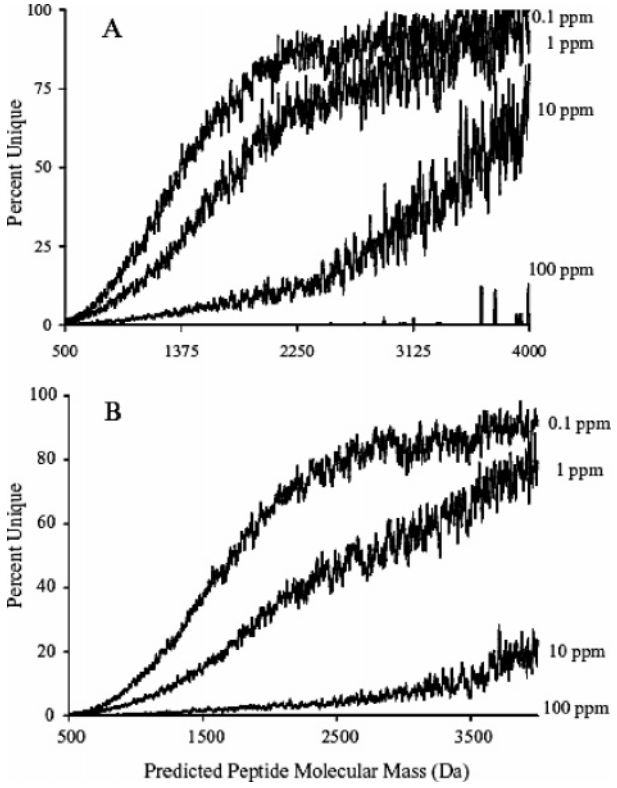


Figure 9. Calculated percent of unique tryptic fragments (potential accurate mass tags) as a function of tryptic fragment mass at four different levels of mass measurement accuracy for the predicted proteins of yeast (A) and *C. elegans* (B). (Reprinted with permission from ref 51. Copyright 2000 American Chemical Society.)

peptide identifications in sub-proteome analyses, due either to the large mass defect of the modified peptides (e.g., phosphopeptides) or to the effectively reduced sample complexity (e.g., cysteinyl peptides).[51] In a study where cysteine-containing peptides were detected at 1 ppm mass accuracy within a peptide mixture by incorporating chlorine into a general alkylation reagent specific for cysteine residues

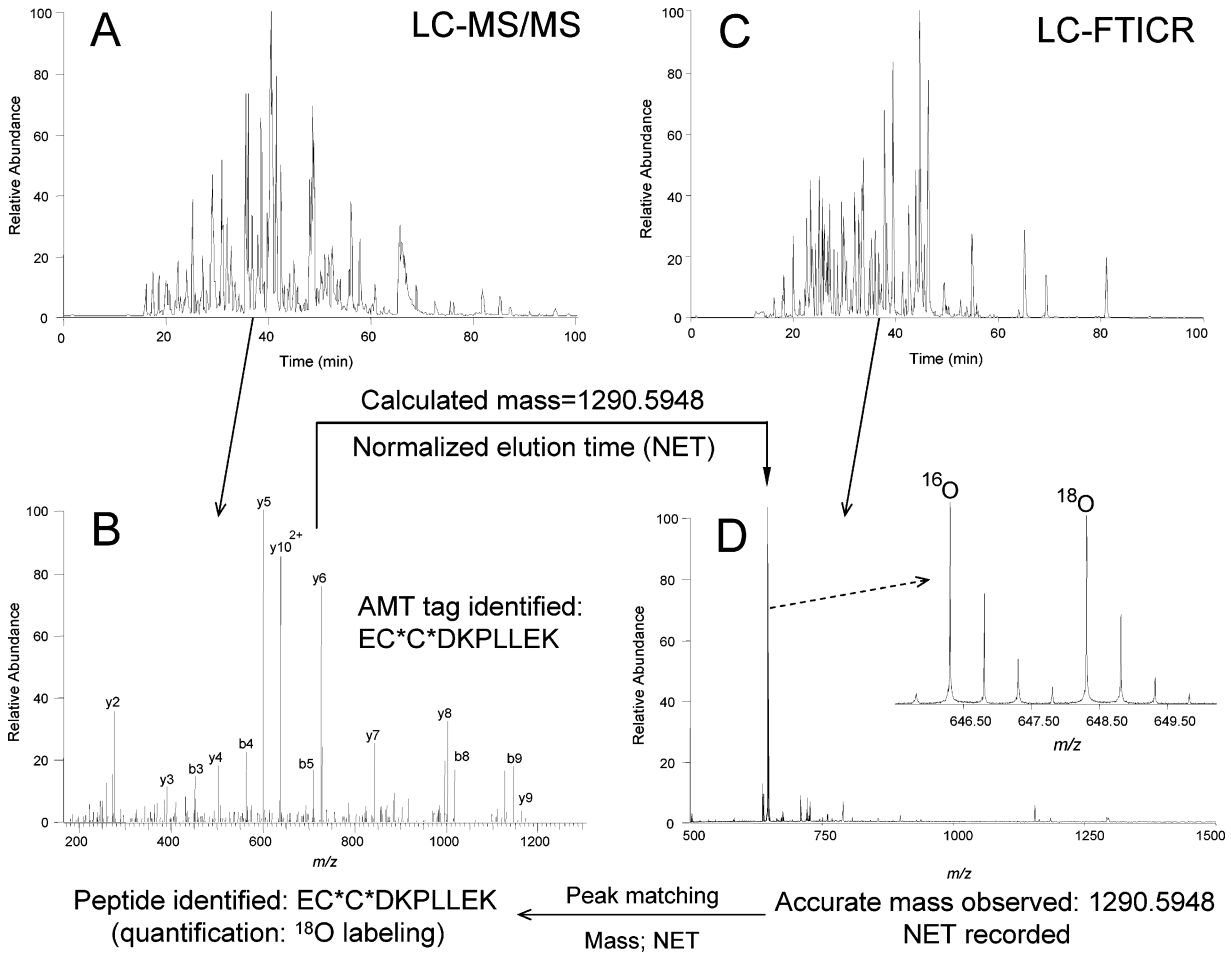


Figure 10. Experimental steps involved in establishing and using an AMT tag. (A) A tryptic digest of a protein mixture is analyzed by LC-MS/MS. (B) A tryptic peptide EC*C*DKPLLEK (C* represents alkylated cysteine residues) is identified by MS/MS. The calculated mass of this peptide (i.e., 1290.5948 Da) and its normalized elution time (NET) are then used to define this peptide in the AMT tag database. (C) In the second stage, the sample is analyzed under the same LC conditions using a FTICR mass spectrometer. (D) The accurate mass (i.e., 1290.5948 Da) and NET observed for a doubly charged peptide are used to match to those of the AMT tags in the database, which leads to its confident identification (EC*C*DKPLLEK). Peptides in isotopically labeled (e.g., ^{18}O labeling) samples can be quantified using the maximum intensities of paired monoisotopic peaks (inset).

(to introduce a mass defect), rapid and unambiguous protein identification could be made by using a single accurate mass of a cysteine-containing peptide and constrained database searching.[215] However, more specificity of the analysis is needed for studying more complex biological systems.

3.3.1. LC-MS Feature Based Profiling for High-Throughput Proteomics

The limited resolving power of MS-only measurements has been largely overcome by utilizing accurate LC retention time information obtained from capillary reversed phase nanoLC separations in addition to accurate mass measurements derived from a high-resolution mass spectrometer such as FTICR to resolve and further identify individual features.[71,321−325] A feature consists of a detected species with an associated unique mass and elution time. Provided the MMA and time measurement accuracy (TMA) are sufficient, each LC-MS feature will be unique among all detectable species from a given biological system. Given a pre-established database of features for a particular biological system, features can be effectively identified based only on their unique mass and time information. An attractive aspect of this approach is that the measured abundances of LC-MS features can be utilized for relative quantitation among different conditions. Statistical analyses can be applied for comparing different conditions for biological characterization or, for example, biomarker discovery, and features of interest can be further subjected to targeted MS/MS analysis if they are not contained within the pre-established database.

This LC-MS feature based peptide/protein identification approach has been initially termed the AMT tag strategy.[71,321,322,324] The first step of this strategy is to establish an extensive LC-MS feature database. Tryptic digests of complex protein mixtures are analyzed using multidimensional LC-MS/MS, and the identified peptides along with their calculated masses and accurate measured elution times are incorporated as AMT tags into the database. As discussed earlier in this paper (section 3.2.1), the use of high-mass-accuracy mass spectrometers (e.g., LTQ-FT or LTQ-Orbitrap) in LC-MS/MS analyses would result in increased confidence in peptide identification and thus improvements in finding/defining the AMT tags. The AMT tag database provides comprehensive coverage of the proteome and serves as a "look-up table" for all subsequent LC-MS proteome analyses without the need for repeated and time-consuming LC-MS/MS analyses of every sample. The experimental steps involved in establishing and using an AMT tag are illustrated in Figure 10. A detected LC-MS feature can be confidently identified when it matches the same elution time and theoretical mass of an AMT tag in the database. The power

of this high-throughput AMT tag approach using high-mass-accuracy data from FTICR and high-resolution separations for effective peptide/protein identifications and quantitation has been demonstrated in a number of applications ranging from microbial to mammalian systems.[324,326−328] Alternatively, experimental and theoretical peptide p*I* information[329] may also be used along with accurate mass information for peptide/protein identification.

The effectiveness of LC-MS feature or AMT tag based identification/quantitation depends on (1) the complexity of the system, (2) the complexity of the detectable species from the analysis of the system, and (3) the overall resolution and accuracy in the mass and time dimensions for LC-MS. The MMA is dependent not only on the instrument resolution but also on other factors, such as calibration and mass shift correction, as previously discussed. The TMA is dependent on the resolution and reproducibility of the LC separation and the accuracy of retention time normalization among analyses. Recent developments in peptide retention time prediction models in reversed phase LC have allowed accurate normalization of retention times between datasets of similar samples by using a genetic linear algorithm (GA) approach.[330,331] It has been recently demonstrated that 1−3 ppm in MMA and 1−2% in TMA for normalized elution time (NET) can be routinely obtained using LC-FTICR.[332] In turn, a distinguishing power equivalent to that achievable using MS alone with a MMA of 0.1 ppm or less is obtained. However, the specificity of the LC-MS measurement is greater than that with MS alone because the specificity in LC-MS reflects both peptide chemical composition and physicochemical properties and, thus, can distinguish a peptide among many (e.g., sequence variants) that have identical masses. With this level of specificity, high-confidence (e.g., FDR of ∼3%) and comprehensive identifications of LC-MS features for even very complex mammalian proteomes, such as human blood plasma[327] and mouse brain tissue,[333] can be made. As shown in Figure 11, the NET constraint significantly reduces the level of random matches, as indicated by the background level for each

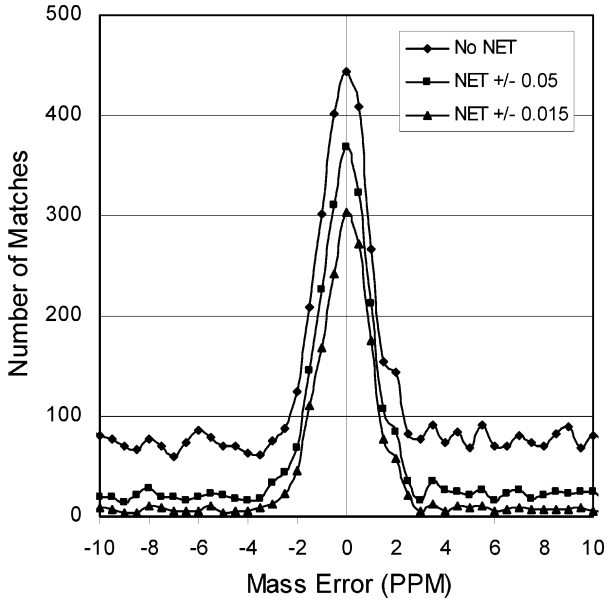


Figure 11. Mass error histograms of features detected from a single LC-FTICR dataset of a human plasma sample that matched to a human plasma AMT tag database using different levels of normalized elution time (NET) constraints. The LC separation time is normalized to a 0−1 scale in NET.

histogram of mass error (the difference between observed mass and calculated mass for the matched peptide in the database) for a human plasma dataset analyzed by LC-FTICR.[332]

Similar to the concept of the AMT tag strategy, a number of other approaches have reported the use of mass and elution time information of high-resolution LC-MS features for comparative proteomic analyses and protein identifications.[325,334−337] In addition, a number of software tools and algorithms including VIPER,[324] msInspect,[338] MapQuant,[339] LCMSWARP,[340] and XCMS[341] have been developed for aligning and normalizing LC-MS features across multiple datasets or experiments. Compared to data-dependent MS/MS approaches, the high-resolution LC-MS feature based approaches have the advantage of high sensitivity and overall proteome coverage due to elimination of the stochastic nature of MS/MS sampling (or undersampling issue) on the chromatographic time scale. As a result of this improved sensitivity and coverage, typically limited or no fractionation is required for high-resolution LC-MS feature based approaches, which increases analytical throughput and allows a larger number of clinical or biological samples to be analyzed for a given study. Additionally, the high-resolution LC-MS feature based approach can be extended to metabolomics applications, such as metabolite profiling in biomarker discovery.

3.3.2. LC-MS Feature Based Quantitative Proteomics

The ability to quantitatively determine changes in protein abundances as well as in protein PTMs in cells, tissues, and biofluids is essential for elucidating cellular processes and signaling pathways and discovering useful candidate protein biomarkers indicative of diseases. When coupled with stable-isotope labeling and "label-free" quantitative approaches, LC-MS feature based profiling is currently the most promising technique for large scale clinical proteomics and protein biomarker discovery applications. To date, stable isotope labeling is still the most commonly used approach for quantitative proteomics, and many different isotope labeling chemistries have been reported.[8,342] In principle, all current quantitative approaches can be easily coupled with LC-MS feature profiling, with the exception of the isobaric tagging approach, which relies on MS/MS fragments for quantitation.[343]

Isotope labeling approaches can generally be divided into three categories: (1) metabolic labeling of proteins by culturing of cells in isotopically enriched media (i.e., enriched with ^{15}N salt, or $^{13}C/^{15}N$ labeled amino acids) or isotopically depleted media;[344,345] (2) enzymatic labeling, such as trypsin-catalyzed oxygen exchange;[327,346−349] and (3) specific isotope-code tagging at certain function groups for either the global proteome or different subproteomes.[45−47,301,302,304,315,350,351] To differentially compare two different samples, one sample is generally labeled with a heavy isotope while the other is labeled with a light isotope. Because the labeled peptide pair is essentially the same chemical species, they coelute during chromatographic separation and have the exact same ionization efficiency, which enables accurate quantitation. The paired species can be determined by certain mass difference, and peptide or protein abundance ratios can be accurately determined by taking the ratio of the MS intensities for the two peptide versions. High resolution and MMA are required for quantitative analysis using isotope labels that have relatively small mass differences between the light and heavy

forms, such as with ^{18}O-labeling that yields a mass difference of 4 Da.

Alternatively, "label-free" direct quantitation approaches are useful because they provide greater flexibility for comparative analyses and simpler sample processing procedures than labeling approaches. Several initial studies suggest that normalized LC-MS peak intensities for detected peptides can be used to compare relative abundances between similar complex samples.[335,352,353] These studies have demonstrated that abundance ratios of separate model proteins may be predicted to within ~20% in complex proteome digests by using measured peptide ion intensities obtained in LC-MS analyses. Among the main challenges for label-free quantitation are multiple issues that affect the usefulness of peptide peak intensities for relative quantitation, such as differences in electrospray ionization efficiencies among different peptides and different samples,[354] differences in the amount of sample injected in each analysis, and sample preparation and instrument reproducibility. These issues are often peptide-dependent, leading to observed disparity among relative abundances of different peptides that originate from the same protein. Improved MMA would effectively reduce the FDR of the analysis and thus help to alleviate, but not eliminate, these issues.

3.4. Intact Protein Analysis

The use of a bottom-up strategy in proteomics relies on the conversion of proteins to peptides via enzymatic digestion prior to MS analysis. The resulting peptide mixture is much higher in complexity but much smaller in size than the original protein mixture; thus, it has a significant advantage of being able to benefit from high MMA and routine low-energy CID for confident identification. Proteomics technologies in the areas of sample preparation, separation, MS analysis, and bioinformatics have advanced and matured to a point where proteomics labs adapting a bottom-up strategy can routinely generate their own modest sets of peptide and protein identifications without the need for advanced and often expensive mass spectrometers. However, a number of limitations inherent with the bottom-up strategy hamper its ability to provide a more comprehensive survey of biological systems. The most apparent obstacle in bottom-up proteomics analyses is that complete sequence coverage of proteins is rarely achieved, especially in the case of global and large-scale proteome analyses. As a result, important information with respect to the native proteins, such as site-specific mutations and PTMs that are often critical to understanding protein function and regulation, may be lost and cannot be examined in a full spectrum. Moreover, attributing certain peptide identifications to a specific protein is often challenging because of the presence of highly homologous proteins and protein isoforms in proteome samples, which also hinders accurate quantitation. In complex organisms, alternative splicing can lead to a significantly increased protein repertoire; for example, up to three-fourths of human genes have at least one variant.[355−357] Therefore, the use of short peptides as proxy markers for genes is inadequate and often misleading.[358] Given these limitations of a bottom-up strategy, approaching analysis from the top-down, i.e., analyzing individual proteins directly by MS or MS/MS, has been increasingly pursued.[91,359,360]

Intact protein analysis has been carried out using ESI-MS/MS in a TQ instrument,[361] in-source decay of ions in MALDI-TOF,[362] and charge reduction of fragment ions from electrosprayed proteins in an IT.[363] Advanced Q-TOF instruments with the help of internal mass calibration can achieve relatively high MMA; for example, human hemoglobin variants that differed by <6 Da (β-chain) were able to be distinguished from normal hemoglobin in heterozygotes by using Q-TOF and the α-chain as internal standard.[364] Both high MMA and mass resolution greatly enhance the confident assignment of protein identity based on molecular mass and the often complicated fragmentation patterns. FTICR has the highest possible mass resolving power (>400 000) and MMA (<1 ppm), as well as the unique capability of fragmenting intact proteins with a variety of techniques. In addition, FTICR is capable of measuring protein molecules as large as 112 508 Da (measured 112 509 Da) at a resolving power of 170 000, using a 9.4 T instrument and a time domain sampling technique.[365] Expectedly, with the use of stronger magnetic fields (e.g., 14.5 T[366]) and ^{13}C and ^{15}N double depletion,[367] even larger proteins can be accurately measured. Another advantage of FTICR is that proteins present at zeptomole[368] to attomole[369] concentrations can be detected, even those in complex protein mixtures.[370,371] Given these advantages, most intact protein analyses have been carried out using FTICR.

3.4.1. Intact Protein Profiling

2-DE has been an important technique in proteomics due to its ability to separate and display thousands of expressed proteins. Some useful information of proteins, such as p*I* and M_r, can also be obtained using this technique. However, 2-DE is relatively low-throughput, is labor-intensive, lacks sensitivity, and still requires subsequent efforts (e.g., in-gel digestion, MS analysis) for protein identification and characterization. New techniques developed to address these limitations typically involve high-performance separations coupled to a highly sensitive and accurate MS instrument (e.g., FTICR). Upfront separation of a protein mixture prior to MS analysis reduces sample complexity and provides useful information about the native protein, such as p*I*, M_r, hydrophobicity, and electrophoretic mobility, depending on the type of separation technique. For instance, capillary isoelectric focusing (CIEF) coupled to FTICR provides p*I* and molecular mass information (analogous to 2-DE) with high resolution and accuracy on both axes, particularly the mass dimension, as well as high sensitivity and throughput. The use of isotope depletion further improved the sensitivity and accuracy of molecular mass measurement in FTICR analyses, as well as significantly enhanced spectral quality.[372] However, even for a simple organism such as *E. coli* and *D. radiodurans*, p*I* and accurate molecular mass alone may still be insufficient in many cases for identifying proteins; additional structural information may need to be acquired via MS/MS on the FTICR instrument.[372,373]

High-pressure (e.g., >10 000 psi) reversed phase LC using a capillary column packed with small particles[374−376] can provide improved recovery for protein separation[377] in addition to providing high peak capacity (~1000) for peptide separations.[378] This technique was coupled to FTICR to characterize intact proteins from the large unit of the yeast ribosome,[136] which was previously complicated due to involvement of a large number of proteins that contained highly basic amino acids and, more significantly, various modifications that often presented in combination. In a single reversed phase LC-FTICR experiment, the high-resolution separation and the high MMA obtained by using "mass

locking" allowed unambiguous identification of 42 of the 43 proteins associated with the core large ribosome subunit and 58 (out of 64 possible) core large ribosomal subunit protein isoforms. This study also demonstrated the significance of intact protein analysis for providing information on cotranslational and post-translational modifications of these ribosomal proteins, which could very well be missed in bottom-up analysis given their highly basic amino acid content.

Considering issues such as different combinations of PTMs and unexpected modifications, it is generally impractical to unambiguously identify proteins at the proteome level based solely on accurate molecular masses. In response to these issues, a strategy that utilizes accurate molecular mass measurements and partial amino acid content information to unambiguously identify intact proteins from sequence databases was developed.[370,371] Proteins were extracted from organisms grown in natural isotopic abundance minimal medium or minimal medium that contained isotopically labeled amino acids (e.g., Leu-D_{10}), after which they were mixed and analyzed by CIEF-FTICR. The accurately measured molecular mass and additional constraint provided by the number of labeled amino acid residues determined from the mass difference of the unlabeled and labeled proteins (Figure 12) facilitates unambiguous protein identification without the need for MS/MS analysis. While protein identification relies more on the amino acid content than on the accurate molecular mass of the unlabeled protein, high MMA greatly aids in the identification of protein PTMs. Simple PTMs, such as the loss of an initiating Met residue, methylation, acetylation, and phosphorylation, can be readily identified by this approach. Identification of more extensive PTMs such as glycosylation may be possible if the number of two or more amino acids present in the protein can be determined and the heterogeneity is not excessive. An apparent limitation of this approach is that only auxotrophic organisms (such as *E. coli* and *S. cerevisiae*) are suitable for this type of study since the labeled amino acid needs to be effectively incorporated by the organism.

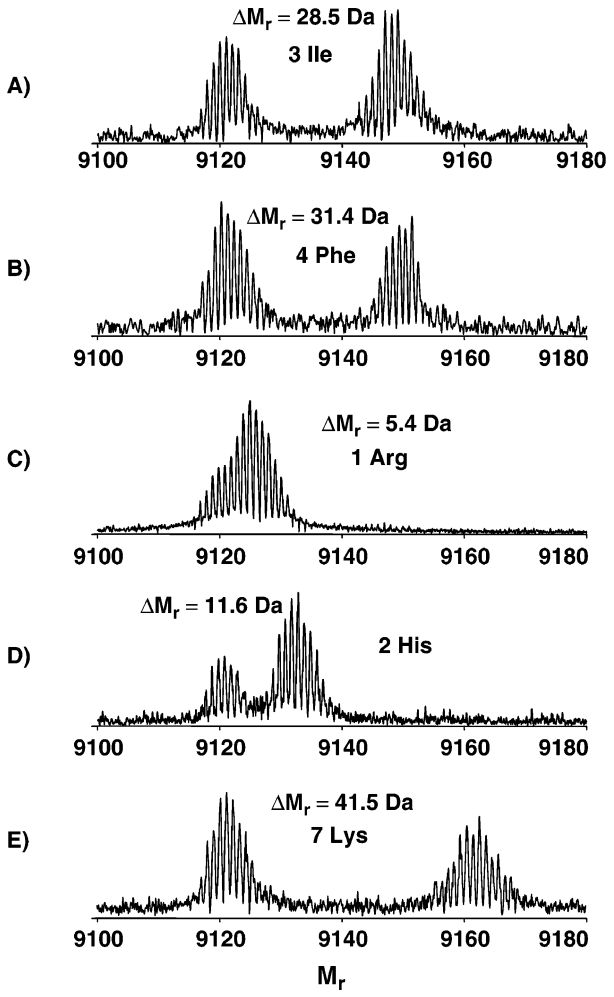


Figure 12. Zero charge state spectra of the *E. coli* phosphotransferase system phosphocarrier protein HPr (M_r = 9119.4 Da) detected during on-line CIEF/FTICR analysis from *E. coli* grown in minimal medium combined with cells grown in minimal medium containing 0.1 mg/mL of (A) Ile-D_{10}, (B) Phe-D_8, (C) Arg-$^{13}C_6$, (D) His-$^{13}C_6$, or (E) Lys-$^{13}C_6$. (Reprinted with permission from ref 371. Copyright 2002 John Wiley & Sons Limited.)

3.4.2. Protein Fragmentation and Characterization

A top-down strategy is particularly attractive for characterizing proteins because protein structure can be determined by using various fragmentation techniques with FTICR instrumentation, such as sustained off-resonance irradiation collision-induced dissociation (SORI-CID),[379] infrared multiphoton dissociation (IRMPD),[380] blackbody infrared radiative dissociation (BIRD),[381] and ECD.[40] The high resolving power and high MMA benefit analysis of both the parent ion and fragment ions that result from backbone bond cleavage of the proteins. In addition, molecular and fragmented masses from the intact protein are far more specific for characterizing protein sequences and PTMs than peptide masses derived from the protein. ECD, which is mainly used on FTICR but is also available on IT instruments,[382−384] induces far more unique cleavages through fast nonergodic dissociation of covalent protein backbone bonds[40,385−387] and allows identification of proteins as large as 45 kDa.[388] Unlike CID, which generates mainly *b* and *y* ions, ECD cleaves the amine bonds to yield *c* and *ż* ions, plus cleavages that produce a small amount of *ȧ* and *y* ions; thus, CID and ECD are complementary.[40,389] The N- and C-terminal fragments in ECD spectra can be readily distinguished without extra chemistry if dissociations between the same residue pair yield both a *y* and a *c* or *ż* ion (the mass difference between *b* and *c* ions is −17.03 Da while the mass difference between *y* and *ż* ions is +16.02 Da), which facilitates automated *de novo* sequencing of proteins. Although for instrumentation reasons ECD is currently mainly available on FTICR instruments, its main benefit is the improved sequence coverage and spectral interpretability, with or without accurate mass measurements. However, ECD with high MMA would certainly further enhance the specificity of the analysis and thus enable the accurate sequencing of larger proteins. For example, the complete sequence of ubiquitin (8.6 kDa) can be correctly predicted *de novo* using ECD with high MMA.[273] Because ECD cleaves predominantly backbone bonds, PTMs that are labile under other activation conditions (e.g., CID, IRMPD) can be retained on the fragments, which allows unambiguous localization of PTMs. An inherent disadvantage of ECD is decreased sensitivity, which is generally lower than that obtained using CID, primarily due to the great variety of fragments produced. As a result, multiple spectra must be added together to improve S/N (i.e., a longer acquisition time is needed) for enhanced identification probability. This requirement compromises coupling ECD with high-performance separation techniques and limits the applicability of ECD for large-

scale global proteome analysis. For example, although CE(LC)-FTICR ECD proved successful for analyzing a simple 3-peptide mixture,[334] analysis of a complex mixture is still challenging.[390]

The combination of accurate mass measurements and ECD appears to be particularly suited for structural characterization of individually expressed proteins, and the problem-solving capabilities of this approach have been demonstrated in a couple of detailed top-down protein analyses. For example, a number of proteins involved in the biosynthesis of thiamin, the biosynthesis of Coenzyme A, and the hydroxylation of proline residues in proteins were overexpressed in *E. coli* and characterized using ECD. Results indicated that most of these proteins exhibited a discrepancy between the predicted and identified sequences in the N-terminus; high-MMA FTICR also allowed identification of unexpected disulfide bond formation in viral prolyl 4-hydroxylase.[388] Top-down analysis with ECD is also a powerful approach for characterizing protein family members and variants, a challenging task for bottom-up proteomics due to the high degree of amino acid sequence reservation, as demonstrated in recent studies on the human histone H2A[391] and H2B[392] families. In these studies, a total of twelve H2A gene family members and two variants[391] and a total of seven H2B gene family members[392] were identified using a top-down strategy and ECD. This approach is particularly useful for H2A and H2B versus other human histones such as H3, where the isoforms can be separated chromatographically. A study of carbonic anhydrase using ECD exemplified protein characterization with the largest number of dissociated interresidue bonds.[393] Cleavages were achieved for this protein at 250 of the 258 interresidue locations by minimizing further cleavage of primary fragments and by denaturing the tertiary noncovalent bonding of the molecular ions under a variety of conditions (e.g., different ESI solutions and ion activation and ECD conditions). This extensive information on backbone bond cleavage can limit the PTM to within one residue.[393] In another study using plasma ECD, all 26 possible phosphorylation sites in casein were characterized.[394]

Most top-down proteomics studies that utilize high-MMA and fragmentation techniques have been demonstrated using only a few intact proteins. An initial FTICR MS/MS study showed that a sequence tag containing three to four contiguous amino acid residues and a molecular mass of <2 kDa was sufficient for protein identification from a species-independent database; however, more sequence tags may be required for unambiguous identification of larger proteins.[395] Later, another statistic model demonstrated probability based protein identification without the need for sequence tags, using wild-type proteins extracted from bacteria and the archaea.[396] Only three to four nonadjacent fragment ions (in this case, from IRMPD or SORI-CID) were needed for intact protein identification with >99% confidence from a database of 5000 proteins. This specificity enables searching without restricting protein molecular mass values to a narrow range, which is particularly useful for identifying multiple proteins from a protein mixture fragmented in parallel (two or three proteins can be identified at once).[396] ETD, an ECD-like fragmentation technique which is typically used with widely accessible quadrupole IT instruments,[308,397] can randomly dissociate large peptide and even intact protein cations on a chromatographic time scale for rapid protein identification. With this method, multiply positively charged proteins are allowed to react with fluoranthene radical anions. After electron transfer, the charge-reduced protein ion dissociates through, most likely, the same mechanism as in ECD and generates N-terminal *c* ions and C-terminal *z* ions. The multiply charged fragment ions can be deprotonated in a second ion/ion reaction with the carboxylate anion of benzoic acid through a mechanism of proton-transfer charge reduction (PTR) to produce a simplified spectrum that facilitates interpretation. The fragment ion information (particularly for the 15−40 amino acids at both the N-terminus and the C-terminus of the protein) and the molecular mass information are then used for protein identification through database searching. This approach was applied to characterize histone H3.1 PTMs and to identify a new member of the H2A gene family.[397] In another study of intact proteins from the *E. coli* 70S ribosomal protein complex, 46 of 55 known unique components as well as a number of their modified forms were identified in a single 90 min automated LC-MS/MS experiment, with the data acquisition rate not greatly slower than that used for acquiring CID spectra on tryptic peptide mixtures.[398] Therefore, ETD provides much higher throughput for top-down analysis, as compared to ECD.

Measuring ETD product ions with high MMA is desired to provide better specificity of protein identification and PTM characterization. However, adaptation of the new hybrid instruments which use an IT as an intermediate storage chamber, mass analyzer, or both (e.g., LTQ-FT, LTQ-Orbitrap, Q-TOF) to accommodate ETD has been technically challenging to realize, due mainly to the difficulty in introducing the anions necessary for ETD. Just recently, a dual ion source concept (i.e., one for generating peptide/protein cations and one for generating reagent anions) which requires minimal instrument modification for implementing ETD reactions on hybrid instrumentation has been proposed. Anions generated by an atmospheric pressure chemical ionization (APCI) source have been shown to induce ETD with varying degrees of efficiency.[399,400] Preparation of ETD-inducing anions via ESI has shown a greater degree of success,[401] and this strategy has been extended for the implementation of ETD on a LTQ-Orbitrap mass spectrometer using paused, dual ESI sources to generate discrete ion populations for subsequent ion/ion reaction in the linear IT.[42] ETD product ions are then injected into the orbitrap for high-resolution and high-mass-accuracy measurement (typically within 2 ppm at a resolution of 60 000). Although this approach has fairly long pulsing times and relatively low electron-transfer efficiency, as compared to conventional ETD instrumentation (i.e., IT), its value for top-down analysis was readily apparent. For example, the *c* and *z* ions that were not identifiable using a linear IT could be easily identified from the ETD data acquired on the orbitrap, resulting in increased sequence coverage and higher specificity for protein identification.[42]

Top-down protein sequencing has also been demonstrated for small proteins (10−25 kDa), using MS/MS and MS^3 in an LTQ-Orbitrap instrument. While CID is known to efficiently fragment proteins in ITs, the lack of sufficient resolution of this type of instrument limits its ability to resolve large protein fragment ions and their charge states. The LTQ-Orbitrap has greatly reduced "TOF effects" compared to the LTQ-FT and is capable of achieving high sensitivity (<50 fmol), high MMA (<3 ppm, using the "lock mass" mode of operation), and high resolving power (60 000), which make it suitable for top-down analysis of proteins. High-quality MS, MS/MS, and MS^3 spectra pro-

vided by the LTQ-Orbitrap have allowed identification of unmodified and modified proteins.[402] Although not demonstrated in this study, detection for on-line separation of a complex mixture having a wide range of abundances is likely to give a wider MMA distribution, and new algorithms are needed to fully utilize the MS^3 data and high-MMA information. Also, similar to intact protein identification that uses other fragmentation methods, e.g., IRMPD, ECD, and ETD, this method is currently applicable to small proteins. For proteins above ~70 kDa (roughly where the top-down protein characterization methods become less effective), limited proteolysis (e.g., using Lys-C) may be applied to generate smaller fragments.[403,404] Alternatively, "prefolding dissociation" (PFD) and a method for conformer disruption that involves ESI solution additives can be applied to proteins with masses >200 kDa. Top-down PFD characterization with a 6-T FTICR instrument demonstrated ~70% sequence coverage on the first ~200 residues of each terminus of large proteins.[405] Expectedly, with the use of a stronger magnetic field, stable PTMs (e.g., methylation, acetylation, oxidation, and deamidation) can be characterized for large proteins.

4. Informatics Algorithms and Pipelines for Interpreting and Applying Accurate Mass Information

Complex spectra from high-resolution mass spectrometers require algorithms for automated interpretation because of the nature of the information generated. LC-MS and LC-MS/MS analyses furnish complementary information and need to be interpreted with different algorithms. While LC-MS/MS analyses produce high-confidence identifications through fragmentation spectra, LC-MS experiments provide a comprehensive sampling of ions and, thus, provide better quantitative information because more time is spent in sampling all the ions rather than focusing on a subset of ions for fragmentation. Higher resolution also helps in separation of overlapping signals from peptides of similar mass and thus provides more precise quantitation by better assignment of abundance to individual peptides. As a result, large scale experiments can benefit by incorporating information from multiple types of analyses on multiple types of instruments and using a pipeline of analytical tools. We discuss algorithms for interpretation of high-resolution data in subsection 4.1 and the main analytical pipelines using high mass accuracy in subsection 4.2.

4.1. Analysis Algorithms

In a high-resolution mass spectrometer, peptides and proteins are typically observed as several related peaks that result from isotope combinations of component elements rather than as single peaks. The overall shape of these related peaks is commonly referred to as an isotopic envelope that depends on the chemical composition of the compound, the natural distribution of the isotopes of the elements that make up the compound, and the resolution of the instrument. Moreover, depending on the charge acquired by the compound, the separation of these related peaks changes because ions are samples in *m*/*z* space rather than mass space. Peptides form complicated isotopic envelopes because carbon and sulfur have relatively high percentages of higher isotopes that occur naturally. Several approaches have been developed to model the isotopic profile of chemical compounds.[406−408] These approaches range from the use of polynomial methods to account for the relative abundance of each isotope in each of the elements in the compounds,[408] to the use of precomputed isotopic profiles of multiple copies of individual atoms to calculate isotopic profiles for new chemical formulas,[406] to the use of sophisticated Fourier transform algorithms that perform a convolution of the mass spectra of the isotopes of each of individual element in a compound[407] for creating theoretical profiles. Depending on the dynamic range of the measurement, a complex mass spectrum from an analysis of a complex biological sample such as human plasma can exhibit hundreds to thousands of features, which makes mass spectral analysis a challenging task.

Several algorithms have been developed to analyze a mass spectrum of a complex protein or peptide mixture and find the components that gave rise to the signals observed in the mass spectrum.[339,409−413] These algorithms are applied to high-mass-accuracy data from FTICR and Q-TOF analyses in which the isotopic envelope resolution allows determination of the charge state. Typically, the process of collapsing peaks from the same chemical compound and charge state into one peak is referred to as "deisotoping", while the process of collapsing different charge state signals into one mass is referred to as "deconvolution".[414] Initial efforts at interpreting spectra focused on deconvolution by looking for alternative charge states of the same feature.[414,415] Subsequently, mass spectra were deisotoped by comparing observed isotopic envelopes against theoretical isotopes from "average" molecules.[410,411] THRASH, which is one of the more well-accepted algorithms for deisotoping of a mass spectrum, does so by scanning through the *m*/*z* range and inspecting each significant peak. To deisotope each peak, THRASH identifies its charge state by using a charge detection algorithm. The charge of a peak is detected by autocorrelating the spectrum around a peak against itself (using a hybrid Patterson and Fourier transform) and looking for the shift that causes high autocorrelation values.[416] The charge of the peak is calculated by using the relationship that this shift should be approximately equal to 1.003/charge of the peak. Once a charge is determined, an average empirical formula is guessed by using the average molecular formula from a database. This *averagine* formula and the resolution of the peak are used to determine the approximate theoretical profile for the peak. Fitting the theoretical profile against the observed mass spectrum is used to decide whether the observed signature is real. If it is real, the related peaks are removed by using the theoretical spectra to identify the related peaks. Alternatively, when a low-quality fit is produced from the automatically detected charge state, all charge states are fitted against the observed profile. This process is repeated for every peak in the mass spectrum. Newer algorithms attempt to deisotope mass spectra in the context of a liquid chromatographic separation.[338,339] These algorithms use the elution profile of peptides as extended information to improve the accuracy and speed. In one such approach,[338] peaks in every spectrum are first determined by using wavelet transforms. Peaks with similar *m*/*z* values presented over multiple spectra are grouped together based on the assumption that they represent the same peptides. Isotopic profiles that exhibit an expected LC elution profile are then tested against theoretical distributions, using a Kullback−Leibler distance to compute the distance between observed *m*/*z* value of the peak apex and the theoretical *m*/*z* value. The theoretical distributions are computed using a single parameter truncated Poisson distribution. Another

approach uses image processing algorithms to determine features present in a sample.[339]

The ability to determine peak charge states in a mass spectrum depends on instrumental resolution. Charge states as high as 10−20 can readily be determined in FTICR mass spectra, while charge states higher than 6 are hard to determine in TOF mass spectra. IT mass spectra are notoriously hard, if not impossible, to deisotope because of low resolution. As a result, pipelines that use IT data do not typically perform deisotoping to determine monoisotopic components but instead focus on peak detection in *m*/*z* space using the two-dimensional information available from the repeated acquisition of mass spectra over an LC separation. For example, in one approach, peaks are separated from noise by requiring that a peak has a certain intensity (S/N > 5), has similar *m*/*z* peaks with high intensities in the neighboring ±5 scans, and has another peak with *m*/*z* within its "isotopic range".[337] An alternate sophisticated approach performs peak detection by using a pixelation approach that prioritizes peaks/pixels into levels with progressively lower stringencies.[417] Each level is determined using an $M - N$ rule that specifies the relative S/N, M (compared to the background intensity, C), and the number of scans (N) for which this signal should be seen. Rules are chosen for each level such that the ith level has $2^{i-1}(1000)$ pixels. Pixels from higher levels are merged into lower-level pixels, and original peaks are preserved if pixel overlap occurs by bisecting any merges that take place. Other alternate algorithms have been described that use three-dimensional data to develop matched-filters for different *m*/*z* bins and then apply these filters to remove background noise and identify significant peaks in the data.[341,418] Obviously, the lower resolution of these instruments reduces the accuracy of the results, producing both higher false positives and false negatives in discovering peptides because masses of peptides cannot be correctly determined for multiply charge peptides where isotopic resolution is not sufficient to determine charge. In addition, quantitative information also suffers because peptides of similar masses end up being observed as a single ion with intensity equal to the sum of their individual intensities.

Interpretation of MS/MS spectra requires a different repertoire of algorithms, and several different algorithms and approaches have been described.[31−33,35,37,419,420] As mentioned earlier, two broad approaches are available for interpreting tandem mass spectra: *de novo* sequencing and database searching. The *de novo* sequencing algorithms such as Peaks, GutenTag, and Lutefisk attempt to computationally identify candidate peptide sequences that would give rise to a mass spectrum by looking at the amino acid mass differences between peaks in the spectrum. Alternatively, the database searching algorithms use a database to choose candidate sequences and match suitable candidates against a spectrum to select a set of candidate matches. Candidate selection is performed on the basis of a score obtained from a routine that constructs the theoretical spectrum for a peptide sequence and calculates the similarity between the observed mass spectrum and the theoretical spectrum. Most of the database searching algorithms differ in the way the theoretical spectrum is constructed and the function that is used to compute the similarity scores. SEQUEST, the first such algorithm to be developed, uses the mathematical cross-correlation function to compute this similarity. To do so, the theoretical spectrum is generated using all possible *b* and *y* ions, with each ion having an equal height. In addition, lower-intensity peaks are added for *a* ions, and ammonia and water loss ions are added for relevant amino acids. The theoretical spectrum and the observed spectrum are padded to 4096 points (by zero filling) and cross-correlated. A major limitation of this approach is that it does not take into account observed fragmentation patterns. Experienced mass spectrometrists apply several heuristic rules when validating spectra; however, incorporation of these rules into search algorithms has been slow. In a recent work, a decision tree of rules for fragmentation patterns was learned from a set of curated spectra and incorporated into a web search engine.[234,421] However, most of the research community is still heavily invested in traditional search engines, and incorporation of new software has been slow as well. Of the more accepted tools, X!Tandem[419] is the only algorithm that attempts to incorporate some common rules into the generation of the theoretical spectrum, and it generates a spectrum in which all peaks do not necessarily have the same height. Its scoring scheme uses a hypergeometric function to calculate similarity between the theoretical and observed profiles. Nevertheless, progress continues to be made in modeling fragmentation patterns[220,422] and tools are expected to continue to improve. An additional complication in the interpretation of MS/MS spectra lies with the database matching of modified peptides; it is typically hard to know from the spectrum itself whether it is modified or not (phosphorylation being a notable exception). To allow for the possibility that a spectrum might be modified, it is essential to match the spectrum multiple times with different modification candidates. Doing so increases the search time linearly with the number of modifications for single modification searches and exponentially when multiple modifications are allowed. A recent approach employs a dynamic programming algorithm run on a large dataset to generate a set of candidate modifications by looking for modified peptides after a first pass search has generated an initial candidate set of proteins.[420]

4.2. Analysis Pipelines

Higher-quality information can be extracted in large scale experiments by organizing both the instrumental and data analyses into pipelines that make use of the complementary information available from the different streams of data. Several computational pipelines have recently been developed that attempt to analyze data globally which use elution time information from couple liquid chromatography systems for improving results. Some of these pipelines use data generated on IT mass spectrometers[337,341,417] while others rely on higher-mass-accuracy instruments such as Q-TOF,[341,423,424] new hybrid instruments such as LTQ-FT and LTQ-Orbitrap,[71,325] or a hybrid set of instruments such as FTICR/Q-TOF and IT.[71,325] While a diversity of pipelines have been explored recently, these pipelines share several similar components in the form of algorithms designed to analyze and collate data. Data analysis components for MS data include algorithms that deconvolute spectra where resolution is high enough for isotopic patterns to be observed and algorithms for finding peaks in the case of IT instruments where isotopic resolution is not available. The deconvolution or the peak picking can be performed either on each individual spectrum separately[409−411,415] or on the entire set of spectra together, in the context of a liquid chromatographic separation.[325,337−339,341,412,417] Data analysis components for MS/MS data include algorithms that interpret tandem mass

spectra (MS^2 to MS^n).[32,33,35,57,61,419,420] Data integration components include algorithms that correct elution time and MMA variability among multiple experiments.[330,337,340,425−428]

The order in which these pieces are applied has given rise to several alternative pipelines. Earlier pipelines proposed processing of individual datasets followed by alignment and collation of results.[71,424] More recently, pipelines have been proposed in which datasets are first minimally process and aligned and then further processing is performed on the globally aligned datasets.[325,337,417] Different filtering rules and criteria are applied to align and process datasets generated on instruments with differing mass accuracies.

An LC separation has additional implications other than its use in noise determination and deisotoping in processing pipelines. The addition of an LC separation system to the front end of a mass spectrometer provides additional coordinates for characterizing peptides. Each peptide elutes from an LC column over a period of time, rather than instantaneously. The peptide elution pattern from a column is referred to as its elution profile, while the time it takes for a peptide to elute from the column is referred to as its elution time. Several approaches have been reported in the literature that describe the theoretical elution intensity profile of a peptide eluting from a reverse phase column. Recently, peptides were reported to show a Gaussian elution time distribution around their ideal values after retention time normalization of LC experiments.[330,340] The use of this elution time in addition to mass for identifying peptides, which was pioneered by the AMT tag strategy,[71] has gained increasing acceptance and has been incorporated into other pipelines.[325,337,338,412] By using mass and elution time dimensions, the confidence in MS/MS and MS identifications can be increased.[429] However, peptide elution times can suffer experimental biases due to dead volume. Additionally, nonsystematic local drifts can take place during the course of an experiment as a result of minor imperfections in chromatography, but these imperfections can be accounted for because they appear to affect the majority of peptide elution times. Several algorithms have been described for aligning datasets.[324,337,340,417,425,426,428] The earliest algorithms were developed for gas chromatography systems and were used to align chromatograms from different datasets.[425,430] These algorithms aligned chromatograms by breaking them into pieces and then allowing the pieces to expand and contract (warping) such that aligned chromatograms had the highest similarity to each other. The similarity was computed as a function of the similarities of the intensities of overlapping points, and the alignment functions were computed using a dynamic programming algorithm.

Alignment of LC-MS/MS datasets to each other has been accomplished by using a genetic algorithm to calculate and predict ideal normalized peptide elution times.[330] The AMT tag strategy also uses an algorithm to perform a linear alignment between scan numbers of features in an LC-MS analysis and the NET of peptides from LC-MS/MS data sets that were previously aligned to each other. Subsequently, a continuous profile model (CPM) has been used to both align and normalize total ion chromatograms from multiple LC-MS datasets.[426] This approach employs Expectation Maximization to generate an ideal total ion chromatogram (TIC) from observed TICs with the use of a model similar to a Hidden Markov Model that specifies how sections of a chromatogram may expand or contract, and the TIC values may also be enhanced or suppressed. The use of this algorithm is limited by its computational speed and by the fact that it presumes TICs capture of all the information needed for aligning two datasets. Alternatively, regression functions (specifically regression splines[431]) have been used iteratively for alignments.[417] Other alignment algorithms have been developed to extend the warping approach to suit the needs of a particular pipeline. These algorithms differ on how experiment sectioning is performed and the scoring scheme used. A method was developed that uses raw data from experiments in computing similarity scores across every pair of spectra in two datasets.[337] This similarity score represents a measure of the relative similarity of the intensity patterns of peaks inside two spectra. The alignment function is only able to move vertically, horizontally, and diagonally from one scan to the other. An extended approach uses a smoothing spline to remove this limitation.[428] A dynamic programming algorithm applicable to the AMT tag approach has also been developed recently.[340] This approach can be applied to processed data in which only individual mass and time features are available rather than entire scans. The algorithm aligns datasets by modeling the variability of mass and elution times of features and by breaking datasets into subsections. A similarity score is computed on subsections of data, and a global alignment is computed without the need for a continuous data profile such as total ion current and raw data from each scan. In addition, this algorithm performs mass recalibration. Other algorithms have also been applied for aligning mass and time pairs. In these algorithms, candidate pairs are first generated on the basis of mass alone, and an initial alignment function is generated and refined iteratively by removing spurious matches.[338,423] One of these approaches starts by first estimating a linear transformation function by robust regression, and it subsequently uses nonlinear smoothing spline regressions on residuals to iteratively improve the fit values.[338]

The relative advantages of different modes of mass spectrometry and different types of instruments have led to the development of three main categories of pipelines, that is, those that use (1) multiple LC-MS/MS experiments with or without quantitative profiling,[337,417] (2) LC-MS based experiments to develop quantitative profiles,[338,341,423] and (3) a hybrid strategy of LC-MS based profiling and LC-MS/MS based identifications.[71,325,335,424] Table 1 summarizes some of the different pipelines currently in use.

When performing quantitative comparisons, MS/MS based strategies use the intensity of the precursor ions in the parent MS scans. A pipeline was reported for IT instruments in which a software suite was used to find features common to multiple LC-MS/MS experiments.[417] Because of the lower resolution of the IT instruments, these datasets were hard to deisotope, and processing was done on peak level information. The software bins peaks from MS scans into m/z bins and uses signal processing algorithms to discover peaks in the chromatographic dimension and to create "pamphlets" that contain pixels for identified peaks. Pamphlets from different experiments are aligned by using a 2-D spline, and peaks from different pamphlets are matched to each other based on their closeness after alignment. The identity of the peaks involved is extracted from the interpretation of the MS/MS spectra related to the peaks and is available provided at least one of the spectra is interpretable. Quantitative information from the precursor ions is used to perform intensity normalization and discover features that change in abundance. Profiling on low-resolution instruments has also

Table 1. Summary of Analysis Pipelines Used in LC-MS Based Proteomics[a]

pipeline	IT	TOF/ Q-TOF	FTICR	LTQ-FT/ LTQ-Orbitrap	deisotoping/peak processing algorithms	alignment algorithms
AMT[71]	×	×	×	×	THRASH	GANET, LCMSWARP
Emili Lab[417]	×				M−N rule pamphlets on peaks	2-D smoothing spline
AMRT[424]		×			MaxEnt, ApexTrac	running median type algorithm
SpecArray[423]		×			PepList	PepArray
Signal Maps[337]	×				signal-to-noise ratio cutoffs	dynamic programming algorithm
XCMS[341]	×	×			MEND	iterative loss
msInspect[338]		×	×	×	elution profile and isotopic fitting (theoetical profile modeled by single parameter truncated Poisson)	robust linear regression and iterative high dimensional

[a] The table summarizes the analysis algorithms related to peak processing and alignment of datasets, and the type of instruments used in recently published pipelines using mass and elution time information to perform abundance profiling on samples. Pipelines using high-mass-resolution instruments use routines to deisotope mass spectra from the isotopic envelopes, while pipelines using lower-mass-resolution instruments perform signal processing on the level of the peaks. The use of alignment algorithms is pervasive across recent pipelines, although the specific method used varies.

been attempted by constructing signal maps.[337] As noted earlier, signal processing algorithms can be used to reduce noise by requiring that peaks be observed over multiple consecutive spectra in the chromatographic dimension. In the reported study,[337] the signal maps were aligned, and similarities and differences were extracted and used for biomarker discovery. Multiple analyses were combined in a progressive strategy of aligning and merging datasets based on similarity. MS/MS identifications were also transferred from low-quality identifications to matching high-quality identifications.

LC-MS based profiling methods are useful for classifying samples, but they provide little information about the identity of important features. The different LC-MS based methods are largely similar and involve detection of features in individual datasets, followed by alignment of multiple datasets to a reference. The next common elements in all the datasets are grouped based on mass and elution time similarity, and a master list is generated that is similar to a peptide array and can be used for profiling and classification of samples.[338,423] Hybrid strategies allow both profiling (from MS level information) and identification by using either databases of identified features[71] or hybrid instruments such as LTQ-FT and LTQ-Orbitrap where identifications from interpretation of concurrent MS/MS spectra can be transferred to features.[325]

5. Conclusions and Outlook

MS has evolved both technologically and conceptually as one of the most important tools in the postgenome era, changing from a means of simply obtaining molecular mass information to a versatile platform for measuring the constituents and dynamics of biological systems. Although developments aimed at improving sensitivity, specificity, and throughput in proteomics are essentially an open-ended endeavor, several trends are currently evident. Increasingly robust biochemical methods are continuously being developed for enriching low-abundance proteins and for isolating specific protein complexes prior to MS analysis. High-resolution separations that employ very small inner diameter (e.g., 10 μm) columns, as well as the use of microfluidics and improved ESI sources (e.g., a multiemitter nanoESI source), can provide significantly improved sensitivity and quantitation. More efficient ion transmission from the source into and through the MS analyzer by means of an electrodynamic ion funnel further enhances analytical sensitivity. The exceptionally high MMA now achievable through the new generation of high-resolution mass spectrometers dramatically enhances the fidelity and robustness of large-scale proteomics analysis with both bottom-up and top-down strategies. The emergence of new hybrid instruments addresses the need for highly accurate yet versatile analyses of proteome samples when higher speed at high resolution is desired. Strategies such as the AMT tag approach and other LC-MS feature based approaches improve throughput and enhance studies designed for probing the dynamics of biological systems.

In parallel with these developments are more robust algorithms and analysis pipelines for accurate interpretation and analysis of the high-quality quantitative MS data essential to proteomics. Continued improvements in data analysis algorithms are required to reduce the FDR of identifications, better deal with the ambiguities in identifications, and increase the true positive rates. Better algorithms for discovering features and "aligning" of datasets continue to be developed as the nature of the data is better understood. In addition, approaches for the analysis of MS/MS fragmentation patterns continue to be studied that will result in better quality identification and higher-confidence results with metrics that characterize this confidence and that will extend proteomics to the broad characterization of PTMs.

The interaction between technology and biology will continue to drive advances in both of these fields, as witnessed in the development and application of proteomics over the past 15 years. As a result of these continuing advances, MS based proteomics will be well-positioned to play an important role in many areas of basic biological research, as well as biomedical research directly associated with human health, such as systems biology,[432,433] and biomarker discovery and validation.[332,434]

6. Abbreviations

2-DE	two-dimensional electrophoresis
ADC	analog-to-digital converter
AGC	automated gain control
AMT	accurate mass and time
APCI	atmospheric pressure chemical ionization
BIRD	blackbody infrared radiative dissociation
CE	capillary electrophoresis
CID	collision-induced dissociation
CIEF	capillary isoelectric focusing
COFI	calibration optimization on fragment ions
CPM	continuous profile model
DE	delayed extraction
DeCAL	deconvolution of Coulombic affected linearity

ECD	electron capture dissociation
ESI	electrospray ionization
EST	expressed sequence tag
ETD	electron-transfer dissociation
FDR	false discovery rate
FTICR	Fourier transform ion cyclotron resonance
fwhm	full width at half-maximum
GA	genetic linear algorithm
HPLC	high-performance LC
IMAC	immobilized metal ion chromatography
IRMPD	infrared multiphoton dissociation
IT	ion trap
LC	liquid chromatography
LC-MS/MS	liquid chromatography coupled to tandem mass spectrometry
m/*z*	mass-to-charge ratio
MALDI	matrix-assisted laser desorption/ionization
MMA	mass measurement accuracy
M_r	molecular weight
MS/MS	tandem mass spectrometry
MS	mass spectrometry
MS^n	multiple MS stage
NET	normalized elution time
oaTOF	orthogonal acceleration TOF
PFD	prefolding dissociation
p*I*	isoelectric point
PMF	peptide mass fingerprinting
ppb	part per billion
ppm	part per million
PSD	postsource decay
PTM	post-translational modification
PTR	proton-transfer charge reduction
QE	quadrupole excitation
Q-TOF	quadrupole TOF
RETOF	reflectron TOF
rms	root-mean-square
S/N	signal-to-noise ratio
SCX	strong cation exchange chromatography
SIM	selected ion monitoring
SORI-CID	sustained off-resonance irradiation CID
SWIFT	stored waveform inverse Fourier transform
TIC	total ion chromatogram
TMA	time measurement accuracy
TOF	time-of-flight
TQ	triple quadrupole

7. Acknowledgments

We thank Dr. Aleksey Tolmachev for helpful comments and critical review of the manuscript. We are also grateful to all past and present members of the Biological Systems Analysis and Mass Spectrometry group at Pacific Northwest National Laboratory (PNNL). We thank the U.S. Department of Energy (DOE) Office of Biological and Environmental Research for long-term research support and FTICR technology development, as well as the National Institutes of Health through the National Center for Research Resources (RR018522) for support of portions of the reviewed research. Our laboratories are located in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the DOE and located at PNNL, which is operated by Battelle Memorial Institute for the DOE under Contract DE-AC05-76RL0 1830.

8. References

(1) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.
(2) Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299.
(3) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151.
(4) Baldwin, M. A. *Methods Enzymol.* **2005**, *402*, 3.
(5) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212.
(6) Aebersold, R.; Goodlett, D. R. *Chem. Rev.* **2001**, *101*, 269.
(7) Aebersold, R. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 685.
(8) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198.
(9) Patterson, S. D.; Aebersold, R. H. *Nat. Genet.* **2003**, *33*, 311.
(10) Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, *70*, 437.
(11) Ferguson, P. L.; Smith, R. D. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 399.
(12) Yates, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 297.
(13) Peng, J. M.; Gygi, S. P. *J. Mass Spectrom.* **2001**, *36*, 1083.
(14) Pandey, A.; Mann, M. *Nature* **2000**, *405*, 837.
(15) Gygi, S. P.; Aebersold, R. *Curr. Opin. Chem. Biol.* **2000**, *4*, 489.
(16) Lambert, J. P.; Ethier, M.; Smith, J. C.; Figeys, D. *Anal. Chem.* **2005**, *77*, 3771.
(17) McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R. *Anal. Chem.* **1997**, *69*, 767.
(18) O'Farrell, P. *J. Biol. Chem.* **1975**, *250*, 4007.
(19) Klose, J. *Humangenetik* **1975**, *26*, 231.
(20) Scheele, G. A. *J. Biol. Chem.* **1975**, *250*, 5375.
(21) Patterson, S. D.; Aebersold, R. *Electrophoresis* **1995**, *16*, 1791.
(22) 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.
(23) James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 58.
(24) Yates, J. R.; Speicher, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397.
(25) Pappin, D. J.; Hojrup, P.; Bleasby, A. J. *Curr. Biol.* **1993**, *3*, 327.
(26) Mann, M.; Hojrup, P.; Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22*, 338.
(27) McLafferty, F. W. *Science* **1981**, *214*, 280.
(28) Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233.
(29) Biemann, K. *Methods Enzymol.* **1990**, *193*, 886.
(30) Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601.
(31) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551.
(32) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976.
(33) Mann, M.; Wilm, M. *Anal. Chem.* **1994**, *66*, 4390.
(34) Qin, J.; Fenyo, D.; Zhao, Y. M.; Hall, W. W.; Chao, D. M.; Wilson, C. J.; Young, R. A.; Chait, B. T. *Anal. Chem.* **1997**, *69*, 3995.
(35) Clauser, K. R.; Baker, P.; Burlingame, A. L. *Anal. Chem.* **1999**, *71*, 2871.
(36) Gras, R.; Muller, M. *Curr. Opin. Mol. Ther.* **2001**, *3*, 526.
(37) Sadygov, R. G.; Cociorva, D.; Yates, J. R. *Nat. Methods* **2004**, *1*, 195.
(38) Edman, P.; Begg, G. *Eur. J. Biochem.* **1967**, *1*, 80.
(39) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 806.
(40) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265.
(41) Reid, G. E.; Shang, H.; Hogan, J. M.; Lee, G. U.; McLuckey, S. A. *J. Am. Chem. Soc.* **2002**, *124*, 7353.
(42) McAlister, G. C.; Phanstiel, D.; Good, D. M.; Berggren, W. T.; Coon, J. J. *Anal. Chem.* **2007**, *79*, 3525.
(43) Anderson, N. L.; Anderson, N. G. *Mol. Cell. Proteomics* **2002**, *1*, 845.
(44) Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. *Nat. Biotechnol.* **2002**, *20*, 301.
(45) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. *Nat. Biotechnol.* **2003**, *21*, 660.
(46) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994.
(47) Liu, T.; Qian, W. J.; Strittmatter, E. F.; Camp, D. G., 2nd; Anderson, G. A.; Thrall, B. D.; Smith, R. D. *Anal. Chem.* **2004**, *76*, 5345.
(48) Wolters, D. A.; Washburn, M. P.; Yates, J. R. *Anal. Chem.* **2001**, *73*, 5683.
(49) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nat. Biotechnol.* **2001**, *19*, 242.
(50) Mann, M. *J. Protein Chem.* **1994**, *13*, 506.
(51) Conrads, T. P.; Anderson, G. A.; Veenstra, T. D.; Pasa-Tolic, L.; Smith, R. D. *Anal. Chem.* **2000**, *72*, 3349.
(52) 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.
(53) Zubarev, R. A.; Hakansson, P.; Sundqvist, B. *Anal. Chem.* **1996**, *68*, 4060.
(54) Berndt, P.; Hobohm, U.; Langen, H. *Electrophoresis* **1999**, *20*, 3521.

(55) Sleno, L.; Volmer, D. A.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 183.
(56) Gorshkov, M. V.; Zubarev, R. A. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3755.
(57) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2002**, *74*, 5383.
(58) Qian, W. J.; Liu, T.; Monroe, M. E.; Strittmatter, E. F.; Jacobs, J. M.; Kangas, L. J.; Petritis, K.; Camp, D. G., 2nd; Smith, R. D. *J. Proteome Res.* **2005**, *4*, 53.
(59) Liu, T.; Qian, W. J.; Gritsenko, M. A.; Xiao, W. Z.; Moldawer, L. L.; Kaushal, A.; Monroe, M. E.; Varnum, S. M.; Moore, R. J.; Purvine, S. O.; Maier, R. V.; Davis, R. W.; Tompkins, R. G.; Camp, D. G., 2nd; Smith, R. D. *Mol. Cell. Proteomics* **2006**, *5*, 1899.
(60) Peng, J. M.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. *J. Proteome Res.* **2003**, *2*, 43.
(61) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2003**, *75*, 4646.
(62) Anderson, D. C.; Li, W.; Payan, D. G.; Noble, W. S. *J. Proteome Res.* **2003**, *2*, 137.
(63) MacCoss, M. J.; Wu, C. C.; Yates, J. R. *Anal. Chem.* **2002**, *74*, 5593.
(64) Moore, R. E.; Young, M. K.; Lee, T. D. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 378.
(65) Eriksson, J.; Fenyo, D. *Proteomics* **2002**, *2*, 262.
(66) Olsen, J. V.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13417.
(67) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J. X.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12130.
(68) Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. *Mol. Cell. Proteomics* **2005**, *4*, 835.
(69) He, F.; Emmett, M. R.; Hakansson, K.; Hendrickson, C. L.; Marshall, A. G. *J. Proteome Res.* **2004**, *3*, 61.
(70) Haas, W.; Faherty, B. K.; Gerber, S. A.; Elias, J. E.; Beausoleil, S. A.; Bakalarski, C. E.; Li, X.; Villen, J.; Gygi, S. P. *Mol. Cell. Proteomics* **2006**, *5*, 1326.
(71) Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y. F.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. *Proteomics* **2002**, *2*, 513.
(72) Zubarev, R.; Mann, M. *Mol. Cell. Proteomics* **2007**, *6*, 377.
(73) Comisarow, M. B.; Marshall, A. G. *J. Chem. Phys.* **1976**, *64*, 110.
(74) Dodonov, A. F.; Chernushevich, I. V.; Dodonova, T. F.; Raznikov, V. V.; Talrose, V. L. Method of Mass-spectrometric Analysis for Time-Of-Flight of Uninterrupted Beam of Ions. U.S.S.R. Patent 1681340, September 30, 1991.
(75) Dawson, J. H. J.; Guilhaus, M. *Rapid Commun. Mass Spectrom.* **1989**, *3*, 155.
(76) Makarov, A. *Anal. Chem.* **2000**, *72*, 1156.
(77) Sommer, H.; Thomas, H. A.; Hipple, J. A. *Phys. Rev.* **1949**, *76*, 1877.
(78) Comisarow, M. B.; Marshall, A. G. *Chem. Phys. Lett.* **1974**, *25*, 282.
(79) Wilkins, C. L.; Chowdhury, A. K.; Nuwaysir, L. M.; Coates, M. L. *Mass Spectrom. Rev.* **1989**, *8*, 67.
(80) Dienes, T.; Pastor, S. J.; Schurch, S.; Scott, J. R.; Yao, J.; Cui, S. L.; Wilkins, C. L. *Mass Spectrom. Rev.* **1996**, *15*, 163.
(81) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1.
(82) Marshall, A. G. *Int. J. Mass Spectrom.* **2000**, *200*, 331.
(83) Marshall, A. G.; Hendrickson, C. L. *Int. J. Mass Spectrom.* **2002**, *215*, 59.
(84) Zhang, L. K.; Rempel, D.; Pramanik, B. N.; Gross, M. L. *Mass Spectrom. Rev.* **2005**, *24*, 286.
(85) Holliman, C. L.; Rempel, D. L.; Gross, M. L. *NATO ASI Ser., Series C: Math. Phys. Sci.* **1996**, *475*, 147.
(86) Amster, I. J. *J. Mass Spectrom.* **1996**, *31*, 1325.
(87) Wilkins, C. L. *Trends Anal. Chem.* **1994**, *13* (Special Issue: Fourier Transform Mass Spectrometry), 223.
(88) Marshall, A. G. *Int. J. Mass Spectrom. Ion Processes* **1996**, *157/158* (Special Issue: Fourier Transform Ion Cyclotron Resonance Mass Spectrometry), 1.
(89) Buchanan, M. V. *Fourier Transform Mass Spectrometry: Evolution, Innovation, and Applications*; ACS Symposium Series 359; Oxford University Press: New York, 1987.
(90) Marshall, A. G.; Verdun, F. R. *Fourier Transforms in NMR, Optical and Mass Spectrometry: A User's Handbook*; Elsevier: Amsterdam, 1990.
(91) Bogdanov, B.; Smith, R. D. *Mass Spectrom. Rev.* **2005**, *24*, 168.
(92) Page, J. S.; Masselon, C. D.; Smith, R. D. *Curr. Opin. Biotechnol.* **2004**, *15*, 3.
(93) Shi, S. D. H.; Hendrickson, C. L.; Marshall, A. G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11532.
(94) Marshall, A. G.; Wang, T. C. L.; Ricca, T. L. *J. Am. Chem. Soc.* **1985**, *107*, 7893.
(95) Shi, S. D. H.; Drader, J. J.; Freitas, M. A.; Hendrickson, C. L.; Marshall, A. G. *Int. J. Mass Spectrom.* **2000**, *196*, 591.
(96) Rakov, V. S.; Futrell, J. H.; Denisov, E. V.; Nikolaev, E. N. *Eur. J. Mass Spectrom.* **2000**, *6*, 299.
(97) Rodgers, R. P.; White, F. M.; Hendrickson, C. L.; Marshall, A. G.; Andersen, K. V. *Anal. Chem.* **1998**, *70*, 4743.
(98) Peurrung, A. J.; Kouzes, R. T.; Barlow, S. E. *Int. J. Mass Spectrom.* **1996**, *158*, 39.
(99) White, W. D.; Malmberg, J. H.; Driscoll, C. F. *Phys. Rev. Lett.* **1982**, *49*, 1822.
(100) Jeffries, J. B.; Barlow, S. E.; Dunn, G. H. *Int. J. Mass Spectrom. Ion Processes* **1983**, *54*, 169.
(101) Beachamp, J. L.; Armstrong, J. T. *Rev. Sci. Instrum.* **1969**, *40*, 123.
(102) Jackson, J. D. *Classical Electrodynamics*, 2nd ed.; John Wiley and Sons: New York, 1975.
(103) Gorshkov, M. V.; Nikolaev, E. N. *Int. J. Mass Spectrom. Ion Processes* **1993**, *125*, 1.
(104) Schweikhard, L.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 433.
(105) Nikolaev, E. N.; Miluchihin, N. V.; Inoue, M. *Int. J. Mass Spectrom. Ion Processes* **1995**, *148*, 145.
(106) Mitchell, D. W.; Smith, R. D. *Phys. Rev. E* **1995**, *52*, 4366.
(107) Francl, T. J.; Sherman, M. G.; Hunter, R. L.; Locke, M. J.; Bowers, W. D.; McIver, R. T. *Int. J. Mass Spectrom. Ion Processes* **1983**, *54*, 189.
(108) Ledford, E. B.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **1984**, *56*, 2744.
(109) Easterling, M. L.; Mize, T. H.; Amster, I. J. *Anal. Chem.* **1999**, *71*, 624.
(110) Laude, D. A., Jr.; Beu, S. C. *Anal. Chem.* **1989**, *61*, 2422.
(111) Schweikhard, L.; Guan, S. H.; Marshall, A. G. *Int. J. Mass Spectrom. Ion Processes* **1992**, *120*, 71.
(112) Speir, J. P.; Gorman, G. S.; Pitsenberger, C. C.; Turner, C. A.; Wang, P. P.; Amster, I. J. *Anal. Chem.* **1993**, *65*, 1746.
(113) OConnor, P. B.; Speir, J. P.; Wood, T. D.; Chorush, R. A.; Guan, Z. Q.; McLafferty, F. W. *J. Mass Spectrom.* **1996**, *31*, 555.
(114) Bruce, J. E.; Anderson, G. A.; Smith, R. D. *Anal. Chem.* **1996**, *68*, 534.
(115) Bruce, J. E.; Anderson, G. A.; Brands, M. D.; Pasa-Tolic, L.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 416.
(116) Masselon, C.; Tolmachev, A. V.; Anderson, G. A.; Harkewicz, R.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 99.
(117) Wineland, D. J.; Dehmelt, H. G. *J. Appl. Phys.* **1975**, *46*, 919.
(118) Hannis, J.; Muddiman, D. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 876.
(119) Chalmers, M. J.; Quinn, J. P.; Blakney, G. T.; Emmett, M. R.; Mischak, H.; Gaskell, S. J.; Marshall, A. G. *J. Proteome Res.* **2003**, *2*, 373.
(120) Wu, S.; Kaiser, N. K.; Meng, D.; Anderson, G. A.; Zhang, K.; Bruce, J. E. *J. Proteome Res.* **2005**, *4*, 1434.
(121) Kloster, M. B. G.; Hannis, J. C.; Muddiman, D. C.; Farrell, N. *Biochemistry* **1999**, *38*, 14731.
(122) Palmer, M. E.; Clench, M. R.; Tetler, L. W.; Little, D. R. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 256.
(123) Henry, K. D.; Williams, E. R.; Wang, B. H.; McLafferty, F. W.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9075.
(124) Henry, K. D.; Quinn, J. P.; McLafferty, F. W. *J. Am. Chem. Soc.* **1991**, *113*, 5447.
(125) O'Connor, P. B.; Costello, C. E. *Anal. Chem.* **2000**, *72*, 5881.
(126) Brock, A.; Horn, D. M.; Peters, E. C.; Shaw, C. M.; Ericson, C.; Phung, Q. T.; Salomon, A. R. *Anal. Chem.* **2003**, *75*, 3419.
(127) Witt, M.; Fuchser, J.; Baykut, G. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 553.
(128) Flora, J. W.; Hannis, J. C.; Muddiman, D. C. *Anal. Chem.* **2001**, *73*, 1247.
(129) Nepomuceno, A. I.; Muddiman, D. C.; Bergen, H. R.; Craighead, J. R.; Burke, M. J.; Caskey, P. E.; Allan, J. A. *Anal. Chem.* **2003**, *75*, 3411.
(130) Schwartz, J. C.; Zhou, X. G.; Bier, M. E. Method and Apparatus of Increasing Dynamic Range and Sensitivity of A Mass Spectrometer. U.S. Patent 5,572,022, November 5, 1996.
(131) Belov, M. E.; Zhang, R.; Strittmatter, E. F.; Prior, D. C.; Tang, K.; Smith, R. D. *Anal. Chem.* **2003**, *75*, 4195.
(132) Senko, M. W.; Hendrickson, C. L.; Pasatolic, L.; Marto, J. A.; White, F. M.; Guan, S. H.; Marshall, A. G. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1824.
(133) Gorshkov, M. V.; Tolic, L. P.; Udseth, H. R.; Anderson, G. A.; Huang, B. M.; Bruce, J. E.; Prior, D. C.; Hofstadler, S. A.; Tang, L. A.; Chen, L. Z.; Willett, J. A.; Rockwood, A. L.; Sherman, M. S.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 692.
(134) Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F. *J. Proteome Res.* **2004**, *3*, 621.

(135) Peterman, S. M.; Mulholland, J. J. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 168.
(136) Lee, S. W.; Berger, S. J.; Martinovic, S.; Pasa-Tolic, L.; Anderson, G. A.; Shen, Y. F.; Zhao, R.; Smith, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5942.
(137) Bruce, J. E.; Anderson, G. A.; Wen, J.; Harkewicz, R.; Smith, R. D. *Anal. Chem.* **1999**, *71*, 2595.
(138) Tolmachev, A. V.; Monroe, M. E.; Jaitly, N.; Petyuk, V. A.; Adkins, J. N.; Smith, R. D. *Anal. Chem.* **2006**, *78*, 8374.
(139) Caravatti, P.; Allemann, M. *Org. Mass Spectrom.* **1991**, *26*, 514.
(140) Beu, S. C.; Laude, D. A. *Anal. Chem.* **1992**, *64*, 177.
(141) Barlow, S. E.; Tinkle, M. D. *Rev. Sci. Instrum.* **2002**, *73*, 4185.
(142) Kingdon, K. H. *Phys. Rev.* **1923**, *21*, 408.
(143) Knight, R. D. *Appl. Phys. Lett.* **1981**, *38*, 221.
(144) Makarov, A. A. Mass Spectrometer. U.S. Patent 5,886,346, March 23, 1999.
(145) Gall, I. N.; Golikov, Y. K.; Aleksandrov, M. L.; Pechalina, Y. E.; Holin, N. A. Time-Of-Flight Mass Spectrometer. U.S.S.R. Patent 1247973, July 30, 1986.
(146) Hardman, M.; Makarov, A. A. *Anal. Chem.* **2003**, *75*, 1699.
(147) Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 659.
(148) Horning, S.; Makarov, A. A.; Denisov, E.; Wieghaus, A.; Malek, R.; Lange, O.; Senko, M. 53rd ASMS Conference on Mass Spectrometry and Applied Topics, San Antonio, TX, 2005.
(149) Yates, J. R.; Cociorva, D.; Liao, L. J.; Zabrouskov, V. *Anal. Chem.* **2006**, *78*, 493.
(150) Olsen, J. V.; de Godoy, L. M. F.; Li, G. Q.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. *Mol. Cell. Proteomics* **2005**, *4*, 2010.
(151) Makarov, A.; Denisov, E.; Lange, O.; Horning, S. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 977.
(152) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Anal. Chem.* **2006**, *78*, 2113.
(153) Cameron, A. E.; Eggers, D. F. *Rev. Sci. Instrum.* **1948**, *19*, 605.
(154) Wiley, W. C.; McLaren, I. H. *Rev. Sci. Instrum.* **1955**, *26*, 1150.
(155) Colby, S. M.; King, T. B.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 865.
(156) Whittal, R. M.; Li, L. *Anal. Chem.* **1995**, *67*, 1950.
(157) Brown, R. S.; Lennon, J. J. *Anal. Chem.* **1995**, *67*, 1998.
(158) Vestal, M. L.; Juhasz, P.; Martin, S. A. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1044.
(159) Mamyrin, B. A.; Karatev, D. V.; Schmikk, D. V. *Sov. Phys. JETP* **1973**, *37*, 45.
(160) Bergmann, T.; Martin, T. P.; Schaber, H. *Rev. Sci. Instrum.* **1989**, *60*, 792.
(161) Price, D.; Milnes, G. J. *Int. J. Mass Spectrom. Ion Processes* **1990**, *99*, 1.
(162) Deleted in proof.
(163) Boyle, J. G.; Whitehouse, C. M. *Anal. Chem.* **1992**, *64*, 2084.
(164) Mirgorodskaya, O. A.; Shevchenko, A. A.; Chernushevich, I. V.; Dodonov, A. F.; Miroshnikov, A. I. *Anal. Chem.* **1994**, *66*, 99.
(165) Verentchikov, A. N.; Ens, W.; Standing, K. G. *Anal. Chem.* **1994**, *66*, 126.
(166) Lewin, M.; Guilhaus, M.; Wildgoose, J.; Hoyes, J.; Bateman, B. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 609.
(167) Dodonov, A. F.; Kozlovski, V. I.; Soulimenkov, I. V.; Raznikov, V. V.; Loboda, A. V.; Zhen, Z.; Horwath, T.; Wollnik, H. *Eur. J. Mass Spectrom.* **2000**, *6*, 481.
(168) Piyadasa, C. K. G.; Hakansson, P.; Ariyaratne, T. R. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 620.
(169) Cotter, R. J. *Time-Of-Flight Mass Spectrometry: Instrumentation and Application in Biological Research*; ACS Professional Reference Books; American Chemical Society: Washington, DC, 1997.
(170) Wollnik, H. *Mass Spectrom. Rev.* **1993**, *12*, 89.
(171) Guilhaus, M.; Selby, D.; Mlynski, V. *Mass Spectrom. Rev.* **2000**, *19*, 65.
(172) Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1990**, *62*, 1836.
(173) Cotter, R. J. *Anal. Chem.* **1992**, *64*, A1027.
(174) Whittal, R. M.; Schriemer, D. C.; Li, L. *Anal. Chem.* **1997**, *69*, 2734.
(175) Edmondson, R. D.; Russell, D. H. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 995.
(176) Zhou, J.; Ens, W.; Standing, K. G.; Verentchikov, A. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 671.
(177) Krutchinsky, A. N.; Loboda, A. V.; Spicer, V. L.; Dworschak, R.; Ens, W.; Standing, K. G. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 508.
(178) O'Connor, P. B.; Costello, C. E. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1862.
(179) Laiko, V. V.; Baldwin, M. A.; Burlingame, A. L. *Anal. Chem.* **2000**, *72*, 652.
(180) Baykut, G.; Jertz, R.; Witt, M. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1238.
(181) Loboda, A. V.; Krutchinsky, A. N.; Bromirski, M.; Ens, W.; Standing, K. G. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1047.
(182) Loboda, A. V.; Ackloo, S.; Chernushevich, I. V. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2508.
(183) Cao, P.; Moini, M. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 864.
(184) Bahr, U.; Karas, M. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1052.
(185) Jiang, L. F.; Moini, M. *Anal. Chem.* **2000**, *72*, 20.
(186) Eckers, C.; Wolff, J. C.; Haskins, N. J.; Sage, A. B.; Giles, K.; Bateman, R. *Anal. Chem.* **2000**, *72*, 3683.
(187) de Biasi, V.; Haskins, N.; Organ, A.; Bateman, R.; Giles, K.; Jarvis, S. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1165.
(188) Zhou, F.; Shui, W. Q.; Lu, Y.; Yang, P. Y.; Gu, Y. L. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 505.
(189) Tyler, A. N.; Clayton, E.; Green, B. N. *Anal. Chem.* **1996**, *68*, 3561.
(190) Blom, K. F. *Anal. Chem.* **2001**, *73*, 715.
(191) Kofeler, H. C.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 406.
(192) Colombo, M.; Sirtori, F. R.; Rizzo, V. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 511.
(193) Strittmatter, E. F.; Rodriguez, N.; Smith, R. D. *Anal. Chem.* **2003**, *75*, 460.
(194) Boguski, M. S.; Lowe, T. M. J.; Tolstoshev, C. M. *Nat. Genet.* **1993**, *4*, 332.
(195) Tonella, L.; Walsh, B. J.; Sanchez, J. C.; Ou, K. L.; 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.
(196) Oconnell, K. L.; Stults, J. T. *Electrophoresis* **1997**, *18*, 349.
(197) Joubert, R.; Strub, J. M.; Zugmeyer, S.; Kobi, D.; Carte, N.; Van, Dorsselaer, A.; Boucherie, H.; Jaquet-Gutfreund, L. *Electrophoresis* **2001**, *22*, 2969.
(198) Langen, H.; Berndt, P.; Roder, D.; Cairns, N.; Lubec, G.; Fountoulakis, M. *Electrophoresis* **1999**, *20*, 907.
(199) 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.
(200) Jungblut, P.; Thiede, B. *Mass Spectrom. Rev.* **1997**, *16*, 145.
(201) Gras, R.; Muller, M.; Gasteiger, E.; Gay, S.; Binz, P. A.; Bienvenut, W.; Hoogland, C.; Sanchez, J. C.; Bairoch, A.; Hochstrasser, D. F.; Appel, R. D. *Electrophoresis* **1999**, *20*, 3535.
(202) Samuelsson, J.; Dalevi, D.; Levander, F.; Rognvaldsson, T. *Bioinformatics* **2004**, *20*, 3628.
(203) Breen, E. J.; Hopwood, F. G.; Williams, K. L.; Wilkins, M. R. *Electrophoresis* **2000**, *21*, 2243.
(204) Schmidt, F.; Schmid, M.; Jungblut, P. R.; Mattow, J.; Facius, A.; Pleissner, K. P. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 943.
(205) Rognvaldsson, T.; Hakkinen, J.; Lindberg, C.; Marko-Varga, G.; Potthast, F.; Samuelsson, J. *J. Chromatogr., B* **2004**, *807*, 209.
(206) Levander, F.; Rognvaldsson, T.; Samuelsson, J.; James, P. *Proteomics* **2004**, *4*, 2594.
(207) Dodds, E. D.; An, H. J.; Hagerman, P. J.; Lebrilla, C. B. *J. Proteome Res.* **2006**, *5*, 1195.
(208) Fenyo, D.; Qin, J.; Chait, B. T. *Electrophoresis* **1998**, *19*, 998.
(209) Jensen, O. N.; Podtelejnikov, A.; Mann, M. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1371.
(210) Beavis, R. C.; Chait, B. T. *Chem. Phys. Lett.* **1991**, *181*, 479.
(211) Pan, Y.; Cotter, R. J. *Org. Mass Spectrom.* **1992**, *27*, 3.
(212) Wong, R. L.; Amster, I. J. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 205.
(213) Eriksson, J.; Chait, B. T.; Fenyo, D. *Anal. Chem.* **2000**, *72*, 999.
(214) Sechi, S.; Chait, B. T. *Anal. Chem.* **1998**, *70*, 5150.
(215) 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.
(216) Hernandez, H.; Niehauser, S.; Boltz, S. A.; Gawandi, V.; Phillips, R. S.; Amster, I. J. *Anal. Chem.* **2006**, *78*, 3417.
(217) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. *Nat. Biotechnol.* **1999**, *17*, 676.
(218) Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. *J. Am. Chem. Soc.* **1996**, *118*, 8365.
(219) Tsaprailis, G.; Nair, H.; Somogyi, A.; Wysocki, V. H.; Zhong, W. Q.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. *J. Am. Chem. Soc.* **1999**, *121*, 5142.
(220) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35*, 1399.
(221) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Anal. Chem.* **1993**, *65*, 425.
(222) Schwartz, B. L.; Bursey, M. M. *Biol. Mass Spectrom.* **1992**, *21*, 92.
(223) Breci, L. A.; Tabb, D. L.; Yates, J. R.; Wysocki, V. H. *Anal. Chem.* **2003**, *75*, 1963.

(224) Gu, C. G.; Tsaprailis, G.; Breci, L.; Wysocki, V. H. *Anal. Chem.* **2000**, *72*, 5804.
(225) Yu, W.; Vath, J. E.; Huberty, M. C.; Martin, S. A. *Anal. Chem.* **1993**, *65*, 3015.
(226) Qin, J.; Chait, B. T. *J. Am. Chem. Soc.* **1995**, *117*, 5411.
(227) Tabb, D. L.; Smith, L. L.; Breci, L. A.; Wysocki, V. H.; Lin, D.; Yates, J. R. *Anal. Chem.* **2003**, *75*, 1155.
(228) Qin, J.; Chait, B. T. *Int. J. Mass Spectrom.* **1999**, *191*, 313.
(229) Steen, H.; Mann, M. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 699.
(230) Susin, S. A.; Lorenzo, H. K.; Zamzami, N.; Marzo, I.; Snow, B. E.; Brothers, G. M.; Mangion, J.; Jacotot, E.; Costantini, P.; Loeffler, M.; Larochette, N.; Goodlett, D. R.; Aebersold, R.; Siderovski, D. P.; Penninger, J. M.; Kroemer, G. *Nature* **1999**, *397*, 441.
(231) Davis, M. T.; Lee, T. D. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 1059.
(232) Ducret, A.; Van Oostveen, I.; Eng, J. K.; Yates, J. R.; Aebersold, R. *Protein Sci.* **1998**, *7*, 706.
(233) Havilio, M.; Haddad, Y.; Smilansky, Z. *Anal. Chem.* **2003**, *75*, 435.
(234) Elias, J. E.; Gibbons, F. D.; King, O. D.; Roth, F. P.; Gygi, S. P. *Nat. Biotechnol.* **2004**, *22*, 214.
(235) Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. *J. Mass Spectrom.* **2001**, *36*, 849.
(236) Masselon, C.; Anderson, G. A.; Harkewicz, R.; Bruce, J. E.; Pasa-Tolic, L.; Smith, R. D. *Anal. Chem.* **2000**, *72*, 1918.
(237) Li, L. J.; Masselon, C. D.; Anderson, G. A.; Pasa-Tolic, L.; Lee, S. W.; Shen, Y. F.; Zhao, R.; Lipton, M. S.; Conrads, T. P.; Tolic, N.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 3312.
(238) Baldwin, M. A. *Mol. Cell. Proteomics* **2004**, *3*, 1.
(239) Carr, S.; Aebersold, R.; Baldwin, M.; Burlingame, A.; Clauser, K.; Nesvizhskii, A. *Mol. Cell. Proteomics* **2004**, *3*, 531.
(240) Olsen, J. V.; Ong, S. E.; Mann, M. *Mol. Cell. Proteomics* **2004**, *3*, 608.
(241) Lasonder, E.; Ishihama, Y.; Andersen, J. S.; Vermunt, A. M. W.; Pain, A.; Sauerwein, R. W.; Eling, W. M. C.; Hall, N.; Waters, A. P.; Stunnenberg, H. G.; Mann, M. *Nature* **2002**, *419*, 537.
(242) Williams, J. D.; Flanagan, M.; Lopez, L.; Fischer, S.; Miller, L. A. D. *J. Chromatogr., A* **2003**, *1020*, 11.
(243) Masselon, C.; Pasa-Tolic, L.; Li, L. J.; Anderson, G. A.; Harkewicz, R.; Smith, R. D. *Proteomics* **2003**, *3*, 1279.
(244) Schlosser, A.; Lehmann, W. D. *Proteomics* **2002**, *2*, 524.
(245) Ji, C. J.; Lo, A.; Marcus, S.; Li, L. *J. Proteome Res.* **2006**, *5*, 2567.
(246) Dieguez-Acuna, F. J.; Gerber, S. A.; Kodama, S.; Elias, J. E.; Beausoleil, S. A.; Faustman, D.; Gygi, S. P. *Mol. Cell. Proteomics* **2005**, *4*, 1459.
(247) Denison, C.; Rudner, A. D.; Gerber, S. A.; Bakalarski, C. E.; Moazed, D.; Gygi, S. P. *Mol. Cell. Proteomics* **2005**, *4*, 246.
(248) Pilch, B.; Mann, M. *Genome Biol.* **2006**, *7*, 10.
(249) Everley, P. A.; Bakalarski, C. E.; Elias, J. E.; Waghorne, C. G.; Beausoleil, S. A.; Gerber, S. A.; Faherty, B. K.; Zetter, B. R.; Gygi, S. P. *J. Proteome Res.* **2006**, *5*, 1224.
(250) Chait, B. T.; Wang, R.; Beavis, R. C.; Kent, S. B. H. *Science* **1993**, *262*, 89.
(251) Cagney, G.; Emili, A. *Nat. Biotechnol.* **2002**, *20*, 163.
(252) Keough, T.; Lacey, M. P.; Youngquist, R. S. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2348.
(253) Lindh, I.; Hjelmqvist, L.; Bergman, T.; Sjovall, A.; Griffiths, W. J. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 673.
(254) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Anal. Chem.* **2003**, *75*, 156A.
(255) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015.
(256) Qin, J.; Herring, C. J.; Zhang, X. L. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 209.
(257) Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Anal. Chem.* **2000**, *72*, 4047.
(258) Gu, S.; Pan, S. Q.; Bradbury, E. M.; Chen, X. *Anal. Chem.* **2002**, *74*, 5774.
(259) Snijders, A. P. L.; de Vos, M. G. J.; Wright, P. C. *J. Proteome Res.* **2005**, *4*, 578.
(260) Fu, Q.; Li, L. J. *Anal. Chem.* **2005**, *77*, 7783.
(261) Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. *J. Comput. Biol.* **1999**, *6*, 327.
(262) Chen, T.; Kao, M. Y.; Tepel, M.; Rush, J.; Church, G. M. *J. Comput. Biol.* **2001**, *8*, 325.
(263) Fernandez-de-Cossio, J.; Gonzalez, J.; Satomi, Y.; Shima, T.; Okumura, N.; Besada, V.; Betancourt, L.; Padron, G.; Shimonishi, Y.; Takao, T. *Electrophoresis* **2000**, *21*, 1694.
(264) Ma, B.; Zhang, K. Z.; Hendrie, C.; Liang, C. Z.; Li, M.; Doherty-Kirby, A.; Lajoie, G. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337.
(265) Taylor, J. A.; Johnson, R. S. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1067.
(266) Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Zubarev, R. A. *J. Proteome Res.* **2005**, *4*, 2348.
(267) Frank, A.; Pevzner, P. *Anal. Chem.* **2005**, *77*, 964.
(268) Searle, B. C.; Dasari, S.; Turner, M.; Reddy, A. P.; Choi, D.; Wilmarth, P. A.; McCormack, A. L.; David, L. L.; Nagalla, S. R. *Anal. Chem.* **2004**, *76*, 2220.
(269) Spengler, B. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 703.
(270) Olson, M. T.; Epstein, J. A.; Yergey, A. L. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1041.
(271) Taylor, J. A.; Johnson, R. S. *Anal. Chem.* **2001**, *73*, 2594.
(272) Wielsch, N.; Thomas, H.; Surendranath, V.; Waridel, P.; Frank, A.; Pevzner, P.; Shevchenko, A. *J. Proteome Res.* **2006**, *5*, 2448.
(273) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10313.
(274) Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A. *Mol. Cell. Proteomics* **2005**, *4*, 1180.
(275) Zhang, Z. Q.; McElvain, J. S. *Anal. Chem.* **2000**, *72*, 2337.
(276) Frank, A. M.; Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A.; Pevzner, P. A. *J. Proteome Res.* **2007**, *6*, 114.
(277) Hunter, T. *Cell* **2000**, *100*, 113.
(278) Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. *Mol. Cell. Proteomics* **2006**, *5*, 2384.
(279) Mann, M.; Ong, S. E.; Gronborg, M.; Steen, H.; Jensen, O. N.; Pandey, A. *Trends Biotechnol.* **2002**, *20*, 261.
(280) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21*, 255.
(281) Peng, J. M.; Schwartz, D.; Elias, J. E.; Thoreen, C. C.; Cheng, D. M.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S. P. *Nat. Biotechnol.* **2003**, *21*, 921.
(282) Wohlschlegel, J. A.; Johnson, E. S.; Reed, S. I.; Yates, J. R. *J. Biol. Chem.* **2004**, *279*, 45662.
(283) Wohlschlegel, J. A.; Johnson, E. S.; Reed, S. I.; Yates, J. R. *J. Proteome Res.* **2006**, *5*, 761.
(284) Pevzner, P. A.; Dancik, V.; Tang, C. L. *J. Comput. Biol.* **2000**, *7*, 777.
(285) Tsur, D.; Tanner, S.; Zandi, E.; Bafna, V.; Pevzner, P. A. *Nat. Biotechnol.* **2005**, *23*, 1562.
(286) Pevzner, P. A.; Mulyukov, Z.; Dancik, V.; Tang, C. L. *Genome Res.* **2001**, *11*, 290.
(287) Han, Y.; Ma, B.; Zhang, K. *J. Bioinf. Comput. Biol.* **2005**, *3*, 697.
(288) Bandeira, N.; Tsur, D.; Frank, A.; Pevzner, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6140.
(289) Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A. *Mol. Cell. Proteomics* **2006**, *5*, 935.
(290) Kocher, T.; Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A. *J. Proteome Res.* **2006**, *5*, 659.
(291) Cooper, J. A.; Sefton, B. M.; Hunter, T. *Methods Enzymol.* **1983**, *99*, 387.
(292) DeGnore, J. P.; Qin, J. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 1175.
(293) Zhang, X. L.; Herring, C. J.; Romano, P. R.; Szczepanowska, J.; Brzeska, H.; Hinnebusch, A. G.; Qin, J. *Anal. Chem.* **1998**, *70*, 2050.
(294) Marshall, A. G.; Hendrickson, C. L.; Shi, S. D. H. *Anal. Chem.* **2002**, *74*, 253A.
(295) Bossio, R. E.; Marshall, A. G. *Anal. Chem.* **2002**, *74*, 1674.
(296) Huddleston, M. J.; Annan, R. S.; Bean, M. F.; Carr, S. A. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 710.
(297) Covey, T. R.; Huang, E. C.; Henion, J. D. *Anal. Chem.* **1991**, *63*, 1193.
(298) Wilm, M.; Neubauer, G.; Mann, M. *Anal. Chem.* **1996**, *68*, 527.
(299) Carr, S. A.; Huddleston, M. J.; Annan, R. S. *Anal. Biochem.* **1996**, *239*, 180.
(300) Neubauer, G.; Mann, M. *Anal. Chem.* **1999**, *71*, 235.
(301) Zhou, H. L.; Watts, J. D.; Aebersold, R. *Nat. Biotechnol.* **2001**, *19*, 375.
(302) Oda, Y.; Nagasu, T.; Chait, B. T. *Nat. Biotechnol.* **2001**, *19*, 379.
(303) Goshe, M. B.; Conrads, T. P.; Panisko, E. A.; Angell, N. H.; Veenstra, T. D.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 2578.
(304) Qian, W. J.; Gosche, M. B.; Camp, D. G.; Yu, L. R.; Tang, K. Q.; Smith, R. D. *Anal. Chem.* **2003**, *75*, 5441.
(305) McLachlin, D. T.; Chait, B. T. *Anal. Chem.* **2003**, *75*, 6826.
(306) Beausoleil, S. A.; Villen, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. *Nat. Biotechnol.* **2006**, *24*, 1285.
(307) King, J. B.; Gross, J.; Lovly, C. M.; Rohrs, H.; Piwnica-Worms, H.; Townsend, R. R. *Anal. Chem.* **2006**, *78*, 2171.
(308) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528.
(309) Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4.
(310) Hart, G. W. *Annu. Rev. Biochem.* **1997**, *66*, 315.
(311) Dell, A.; Morris, H. R. *Science* **2001**, *291*, 2351.
(312) Carr, S. A.; Huddleston, M. J.; Bean, M. F. *Protein Sci.* **1993**, *2*, 183.
(313) Huddleston, M. J.; Bean, M. F.; Carr, S. A. *Anal. Chem.* **1993**, *65*, 877.
(314) Jebanathirajah, J.; Steen, H.; Roepstorff, P. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 777.

(315) Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. *Mol. Cell. Proteomics* **2002**, *1*, 791.
(316) Zhang, H.; Yi, E. C.; Li, X. J.; Mallick, P.; Kelly-Spratt, K. S.; Masselon, C. D.; Camp, D. G., 2nd; Smith, R. D.; Kemp, C. J.; Aebersold, R. *Mol. Cell. Proteomics* **2005**, *4*, 144.
(317) Liu, T.; Qian, W. J.; Gritsenko, M. A.; Camp, D. G., 2nd; Monroe, M. E.; Moore, R. J.; Smith, R. D. *J. Proteome Res.* **2005**, *4*, 2070.
(318) Kuster, B.; Mann, M. *Anal. Chem.* **1999**, *71*, 1431.
(319) Medzihradszky, K. F.; Besman, M. J.; Burlingame, A. L. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 472.
(320) Lehmann, W. D.; Bohne, A.; von der Lieth, C. W. *J. Mass Spectrom.* **2000**, *35*, 1335.
(321) Qian, W. J.; Camp, D. G., 2nd; Smith, R. D. *Expert Rev. Proteomics* **2004**, *1*, 87.
(322) Pasa-Tolic, L.; Masselon, C.; Barry, R. C.; Shen, Y. F.; Smith, R. D. *BioTechniques* **2004**, *37*, 621.
(323) Norbeck, A. D.; Monroe, M. E.; Adkins, J. N.; Anderson, K. K.; Daly, D. S.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1239.
(324) Zimmer, J. S. D.; Monroe, M. E.; Qian, W. J.; Smith, R. D. *Mass Spectrom. Rev.* **2006**, *25*, 450.
(325) Jaffe, J. D.; Mani, D. R.; Leptos, K. C.; Church, G. M.; Gillette, M. A.; Carr, S. A. *Mol. Cell. Proteomics* **2006**, *5*, 1927.
(326) Lipton, M. S.; Pasa-Tolic, L.; Anderson, G. A.; Anderson, D. J.; Auberry, D. L.; Battista, K. R.; Daly, M. J.; Fredrickson, J.; Hixson, K. K.; Kostandarithes, H.; Masselon, C.; Markillie, L. M.; Moore, R. J.; Romine, M. F.; Shen, Y. F.; Stritmatter, E.; Tolic, N.; Udseth, H. R.; Venkateswaran, A.; Wong, L. K.; Zhao, R.; Smith, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11049.
(327) Qian, W. J.; Monroe, M. E.; Liu, T.; Jacobs, J. M.; Anderson, G. A.; Shen, Y. F.; Moore, R. J.; Anderson, D. J.; Zhang, R.; Calvano, S. E.; Lowry, S. F.; Xiao, W. Z.; Moldawer, L. L.; Davis, R. W.; Tompkins, R. G.; Camp, D. G., 2nd; Smith, R. D. *Mol. Cell. Proteomics* **2005**, *4*, 700.
(328) Smith, R. D.; Pasa-Tolic, L.; Lipton, M. S.; Jensen, P. K.; Anderson, G. A.; Shen, Y. F.; Conrads, T. P.; Udseth, H. R.; Harkewicz, R.; Belov, M. E.; Masselon, C.; Veenstra, T. D. *Electrophoresis* **2001**, *22*, 1652.
(329) Cargile, B. J.; Talley, D. L.; Stephenson, J. L. *Electrophoresis* **2004**, *25*, 936.
(330) Petritis, K.; Kangas, L. J.; Ferguson, P. L.; Anderson, G. A.; Pasa-Tolic, L.; Lipton, M. S.; Auberry, K. J.; Strittmatter, E. F.; Shen, Y. F.; Zhao, R.; Smith, R. D. *Anal. Chem.* **2003**, *75*, 1039.
(331) Petritis, K.; Kangas, L. J.; Yan, B.; Monroe, M. E.; Strittmatter, E. F.; Qian, W. J.; Adkins, J. N.; Moore, R. J.; Xu, Y.; Lipton, M. S.; Camp, D. G., 2nd; Smith, R. D. *Anal. Chem.* **2006**, *78*, 5026.
(332) Qian, W. J.; Jacobs, J. M.; Liu, T.; Camp, D. G., 2nd; Smith, R. D. *Mol. Cell. Proteomics* **2006**, *5*, 1727.
(333) Petyuk, V. A.; Qian, W. J.; Chin, M. H.; Wang, H. X.; Livesay, E. A.; Monroe, M. E.; Adkins, J. N.; Jaitly, N.; Anderson, D. J.; Camp, D. G.; Smith, D. J.; Smith, R. D. *Genome Res.* **2007**, *17*, 328.
(334) Palmblad, M.; Ramstrom, M.; Markides, K. E.; Hakansson, P.; Bergquist, J. *Anal. Chem.* **2002**, *74*, 5826.
(335) Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C. H. *Anal. Chem.* **2003**, *75*, 4818.
(336) Chen, S. S.; Deutsch, E. W.; Yi, E. C.; Li, X. J.; Goodlettt, D. R.; Aebersold, R. *J. Proteome Res.* **2005**, *4*, 2174.
(337) Prakash, A.; Mallick, P.; Whiteaker, J.; Zhang, H. D.; Paulovich, A.; Flory, M.; Lee, H.; Aebersold, R.; Schwikowski, B. *Mol. Cell. Proteomics* **2006**, *5*, 423.
(338) Bellew, M.; Coram, M.; Fitzgibbon, M.; Igra, M.; Randolph, T.; Wang, P.; May, D.; Eng, J.; Fang, R. H.; Lin, C. W.; Chen, J. Z.; Goodlett, D.; Whiteaker, J.; Paulovich, A.; McIntosh, M. *Bioinformatics* **2006**, *22*, 1902.
(339) Leptos, K. C.; Sarracino, D. A.; Jaffe, J. D.; Krastins, B.; Church, G. M. *Proteomics* **2006**, *6*, 1770.
(340) Jaitly, N.; Monroe, M. E.; Petyuk, V. A.; Clauss, T. R. W.; Adkins, J. N.; Smith, R. D. *Anal. Chem.* **2006**, *78*, 7397.
(341) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779.
(342) Ong, S. E.; Mann, M. *Nat. Chem. Biol.* **2005**, *1*, 252.
(343) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. *Mol. Cell. Proteomics* **2004**, *3*, 1154.
(344) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6591.
(345) 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.
(346) Mirgorodskaya, O. A.; Kozmin, Y. P.; Titov, M. I.; Korner, R.; Sonksen, C. P.; Roepstorff, P. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1226.
(347) Stewart, I.; Thomson, T.; Figeys, D. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2456.
(348) Yao, X. D.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 2836.
(349) Heller, M.; Mattou, H.; Menzel, C.; Yao, X. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 704.
(350) Yu, L. R.; Conrads, T. P.; Uo, T.; Issaq, H. J.; Morrison, R. S.; Veenstra, T. D. *J. Proteome Res.* **2004**, *3*, 469.
(351) Chakraborty, A.; Regnier, F. E. *J. Chromatogr., A* **2002**, *949*, 173.
(352) Chelius, D.; Bondarenko, P. V. *J. Proteome Res.* **2002**, *1*, 317.
(353) Fang, R. H.; Elias, D. A.; Monroe, M. E.; Shen, Y. F.; McIntosh, M.; Wang, P.; Goddard, C. D.; Callister, S. J.; Moore, R. J.; Gorby, Y. A.; Adkins, J. N.; Fredrickson, J. K.; Lipton, M. S.; Smith, R. D. *Mol. Cell. Proteomics* **2006**, *5*, 714.
(354) Smith, R. D.; Shen, Y. F.; Tang, K. Q. *Acc. Chem. Res.* **2004**, *37*, 269.
(355) Johnson, J. M.; Castle, J.; Garrett-Engele, P.; Kan, Z. Y.; Loerch, P. M.; Armour, C. D.; Santos, R.; Schadt, E. E.; Stoughton, R.; Shoemaker, D. D. *Science* **2003**, *302*, 2141.
(356) Stamm, S.; Ben-Ari, S.; Rafalska, I.; Tang, Y. S.; Zhang, Z. Y.; Toiber, D.; Thanaraj, T. A.; Soreq, H. *Gene* **2005**, *344*, 1.
(357) Resch, A.; Xing, Y.; Modrek, B.; Gorlick, M.; Riley, R.; Lee, C. *J. Proteome Res.* **2004**, *3*, 76.
(358) Godovac-Zimmermann, J.; Kleiner, O.; Brown, L. R.; Drukier, A. K. *Proteomics* **2005**, *5*, 699.
(359) Reid, G. E.; McLuckey, S. A. *J. Mass Spectrom.* **2002**, *37*, 663.
(360) Kelleher, N. L. *Anal. Chem.* **2004**, *76*, 196A.
(361) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Science* **1990**, *248*, 201.
(362) Reiber, D. C.; Grover, T. A.; Brown, R. S. *Anal. Chem.* **1998**, *70*, 673.
(363) Cargile, B. J.; McLuckey, S. A.; Stephenson, J. L. *Anal. Chem.* **2001**, *73*, 1277.
(364) Rai, D. K.; Griffiths, W. J.; Landin, B.; Wild, B. J.; Alvelius, G.; Green, B. N. *Anal. Chem.* **2003**, *75*, 1978.
(365) Kelleher, N. L.; Senko, M. W.; Siegel, M. M.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 380.
(366) Marshall, A. G. *Physica B* **2004**, *346*, 503.
(367) Marshall, A. G.; Senko, M. W.; Li, W. Q.; Li, M.; Dillon, S.; Guan, S. H.; Logan, T. M. *J. Am. Chem. Soc.* **1997**, *119*, 433.
(368) Belov, M. E.; Gorshkov, M. V.; Udseth, H. R.; Anderson, G. A.; Smith, R. D. *Anal. Chem.* **2000**, *72*, 2271.
(369) Valaskovic, G. A.; Kelleher, N. L.; McLafferty, F. W. *Science* **1996**, *273*, 1199.
(370) Veenstra, T. D.; Martinovic, S.; Anderson, G. A.; Pasa-Tolic, L.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 78.
(371) Martinovic, S.; Veenstra, T. D.; Anderson, G. A.; Pasa-Tolic, L.; Smith, R. D. *J. Mass Spectrom.* **2002**, *37*, 99.
(372) 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.
(373) Jensen, P. K.; Pasa-Tolic, L.; Anderson, G. A.; Horner, J. A.; Lipton, M. S.; Bruce, J. E.; Smith, R. D. *Anal. Chem.* **1999**, *71*, 2076.
(374) MacNair, J. E.; Lewis, K. C.; Jorgenson, J. W. *Anal. Chem.* **1997**, *69*, 983.
(375) MacNair, J. E.; Patel, K. D.; Jorgenson, J. W. *Anal. Chem.* **1999**, *71*, 700.
(376) Tolley, L.; Jorgenson, J. W.; Moseley, M. A. *Anal. Chem.* **2001**, *73*, 2985.
(377) Eschelbach, J. W.; Jorgenson, J. W. *Anal. Chem.* **2006**, *78*, 1697.
(378) Shen, Y. F.; Tolic, N.; Zhao, R.; Pasa-Tolic, L.; Li, L. J.; Berger, S. J.; Harkewicz, R.; Anderson, G. A.; Belov, M. E.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 3011.
(379) Gauthier, J. W.; Trautman, T. R.; Jacobson, D. B. *Anal. Chim. Acta* **1991**, *246*, 211.
(380) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 2809.
(381) Price, W. D.; Schnier, P. D.; Williams, E. R. *Anal. Chem.* **1996**, *68*, 859.
(382) Baba, T.; Hashimoto, Y.; Hasegawa, H.; Hirabayashi, A.; Waki, I. *Anal. Chem.* **2004**, *76*, 4263.
(383) Silivra, O. A.; Kjeldsen, F.; Ivonin, I. A.; Zubarev, R. A. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 22.
(384) Ding, L.; Brancia, F. L. *Anal. Chem.* **2006**, *78*, 1995.
(385) Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 2857.
(386) 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.

(387) Cerda, B. A.; Horn, D. M.; Breuker, K.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 9287.
(388) Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J. H.; Begley, T. P.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 672.
(389) Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A. *Angew. Chem., Int. Ed.* **2006**, *45*, 5301.
(390) Tsybin, Y. O.; Hakansson, P.; Wetterhall, M.; Markides, K. E.; Bergquist, J. *Eur. J. Mass Spectrom.* **2002**, *8*, 389.
(391) Boyne, M. T.; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. *J. Proteome Res.* **2006**, *5*, 248.
(392) Siuti, N.; Roth, M. J.; Mizzen, C. A.; Kelleher, N. L.; Pesavento, J. J. *J. Proteome Res.* **2006**, *5*, 233.
(393) Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1774.
(394) Sze, S. K.; Ge, Y.; Oh, H. B.; McLafferty, F. W. *Anal. Chem.* **2003**, *75*, 1599.
(395) Mortz, E.; O'Connor, P. B.; Roepstorff, P.; Kelleher, N. L.; Wood, T. D.; McLafferty, F. W.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8264.
(396) Meng, F. Y.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. *Nat. Biotechnol.* **2001**, *19*, 952.
(397) Coon, J. J.; Ueberheide, B.; Syka, J. E. P.; Dryhurst, D. D.; Ausio, J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9463.
(398) Chi, A.; Bai, D. L.; Geer, L. Y.; Shabanowitz, J.; Hunt, D. F. *Int. J. Mass Spectrom.* **2007**, *259*, 197.
(399) Liang, X. R.; Xia, Y.; McLuckey, S. A. *Anal. Chem.* **2006**, *78*, 3208.
(400) Xia, Y.; Chrisman, P. A.; Erickson, D. E.; Liu, J.; Liang, X. R.; Londry, F. A.; Yang, M. J.; McLuckey, S. A. *Anal. Chem.* **2006**, *78*, 4146.
(401) Huang, T. Y.; Emory, J. F.; O'Hair, R. A. J.; McLuckey, S. A. *Anal. Chem.* **2006**, *78*, 7387.
(402) Macek, B.; Waanders, L. F.; Olsen, J. V.; Mann, M. *Mol. Cell. Proteomics* **2006**, *5*, 949.
(403) Forbes, A. J.; Mazur, M. T.; Kelleher, N. L.; Patel, H. M.; Walsh, C. T. *Eur. J. Mass Spectrom.* **2001**, *7*, 81.
(404) Forbes, A. J.; Mazur, M. T.; Patel, H. M.; Walsh, C. T.; Kelleher, N. L. *Proteomics* **2001**, *1*, 927.
(405) Han, X. M.; Jin, M.; Breuker, K.; McLafferty, F. W. *Science* **2006**, *314*, 109.
(406) Kubinyi, H. *Anal. Chim. Acta* **1991**, *247*, 107.
(407) Rockwood, A. L.; Vanorden, S. L.; Smith, R. D. *Anal. Chem.* **1995**, *67*, 2699.
(408) Yergey, J. A. *Int. J. Mass Spectrom. Ion Phys.* **1983**, *52*, 337.
(409) Du, P. C.; Angeletti, R. H. *Anal. Chem.* **2006**, *78*, 3385.
(410) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320.
(411) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229.
(412) Katajamaa, M.; Miettinen, J.; Oresic, M. *Bioinformatics* **2006**, *22*, 634.
(413) Ferrige, A. G.; Seddon, M. J.; Skilling, J.; Ordsmith, N. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 765.
(414) Mann, M.; Meng, C. K.; Fenn, J. B. *Anal. Chem.* **1989**, *61*, 1702.
(415) Zhang, Z. Q.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 225.
(416) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 52.
(417) Radulovic, D.; Jelveh, S.; Ryu, S.; Hamilton, T. G.; Foss, E.; Mao, Y.; Emili, A. *Mol. Cell. Proteomics* **2004**, *3*, 984.
(418) Andreev, V. P.; Rejtar, T.; Chen, H. S.; Moskovets, E. V.; Ivanov, A. R.; Karger, B. L. *Anal. Chem.* **2003**, *75*, 6314.
(419) Craig, R.; Beavis, R. C. *Bioinformatics* **2004**, *20*, 1466.
(420) Tanner, S.; Shu, H.; Frank, A.; Wang, L. C.; Zandi, E.; Mumby, M.; Pevzner, P. A.; Bafna, V. *Anal. Chem.* **2005**, *77*, 4626.
(421) Gibbons, F. D.; Elias, J. E.; Gygi, S. P.; Roth, F. P. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 910.
(422) Zhang, Z. *Anal. Chem.* **2005**, *77*, 6364.
(423) Li, X. J.; Yi, E. C.; Kemp, C. J.; Zhang, H.; Aebersold, R. *Mol. Cell. Proteomics* **2005**, *4*, 1328.
(424) Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G. Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P.; Geromanos, S. *Anal. Chem.* **2005**, *77*, 2187.
(425) Bylund, D.; Danielsson, R.; Malmquist, G.; Markides, K. E. *J. Chromatogr., A* **2002**, *961*, 237.
(426) Listgarten, J.; Neal, R. M.; Roweis, S. T.; Emili, A. *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, 2005.
(427) Pierce, K. M.; Wood, L. F.; Wright, B. W.; Synovec, R. E. *Anal. Chem.* **2005**, *77*, 7735.
(428) Prince, J. T.; Marcotte, E. M. *Anal. Chem.* **2006**, *78*, 6140.
(429) Strittmatter, E. F.; Kangas, L. J.; Petritis, K.; Mottaz, H. M.; Anderson, G. A.; Shen, Y. F.; Jacobs, J. M.; Camp, D. G.; Smith, R. D. *J. Proteome Res.* **2004**, *3*, 760.
(430) Nielsen, N. P. V.; Carstensen, J. M.; Smedsgaard, J. *J. Chromatogr., A* **1998**, *805*, 17.
(431) Hastie, T.; Tibshirani, R.; Friedman, J. H. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*; Springer: New York, 2001.
(432) Hood, L.; Perlmutter, R. M. *Nat. Biotechnol.* **2004**, *22*, 1215.
(433) Hood, L.; Heath, J. R.; Phelps, M. E.; Lin, B. Y. *Science* **2004**, *306*, 640.
(434) Rifai, N.; Gillette, M. A.; Carr, S. A. *Nat. Biotechnol.* **2006**, *24*, 971.

CR068288J