REVIEW

ACS chemical biology

Achieving In-Depth Proteomics Profiling by Mass Spectrometry

Natalie G. Ahn[†,‡,*], John B. Shabb[§], William M. Old[†], and Katheryn A. Resing[†]

[†]Department of Chemistry and Biochemistry, [‡]Howard Hughes Medical Institute, University of Colorado, Boulder, Colorado 80309, [§]Department of Biochemistry and Molecular Biology, University of North Dakota, Grand Forks, North Dakota 58202

ABSTRACT Proteomics addresses the important goal of determining the chemistry and composition of proteins in biological samples. Mass-spectrometry-based strategies have been highly successful in identifying and profiling proteins in complex mixtures; however, although depth of sampling continues to improve, a general recognition exists that no study has yet achieved complete protein coverage in any tissue, cell type, subcellular component, or fluid. The development of new approaches for comprehensively surveying highly complex protein mixtures, distinguishing protein isoforms, quantifying changes in protein abundance between different samples, and mapping post-translational modifications are areas of active research. These will be needed to achieve the "systems-wide" protein profiling goals of defining molecular responses to cell perturbations and obtaining biomarker information for disease detection, prognosis, and responses to therapy. We review recent progress in approaching these problems and present examples of successful applications and the outlook for the future.

Protein profiling confers unique advantages over messenger RNA (mRNA) screens for detecting post-transcriptional and post-translational events, tracking protein turnover and subcellular localization, and analyzing fluids and other samples that lack nucleic acids. Early efforts to profile proteins created databases of protein gel mobilities following separation by 2D gel electrophoresis (2DE) (*1*). This approach was advanced by the introduction of mass spectrometry (MS) instruments capable of complex manipulations of ions in the gas phase and thus allowed fragmentation and sequencing of peptides. Examples of fragmentation processes in commonly used triple quadrupole and ion trap mass spectrometers are shown in Figure 1. In each case, peptide ions within a mixture are introduced into the mass spectrometer, and they are manipulated with electric fields to isolate one peptide analyte. The isolated ion then collides with inert gases such as nitrogen or helium in a process referred to as collisional induced dissociation (CID); this process imparts sufficient energy to break bonds within the peptide, producing many fragment ions. The final step "reads out" these fragment ions to produce an MS/MS spectrum, so named because of the two-stage process. Concentrating a fragment ion and further cleaving it by CID to produce a three-stage MS^3 spectrum are also possible. A recent review outlines the characteristics of different MS instruments (*2*).

Protein chemists exploited these methods, using in-gel digestion of excised gel pieces from 2D gels, followed by matching peptide mass and sequence data against genome database entries to identify proteins (*3*). Subsequent development of public web sites with computational algorithms for searching the protein databases (*e.g.*, http://us.expasy.org/tools (*4*–*6*)) has made protein identification by in-gel digestion and MS a

***Corresponding author,**
natalie.ahn@colorado.edu.

Received for review August 17, 2006 and accepted December 7, 2006.

Published online January 19, 2007

10.1021/cb600357d CCC: $37.00

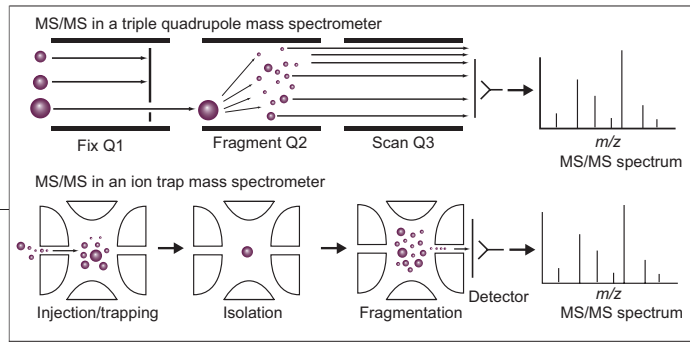

Figure 1. Peptide ion fragmentation by MS/MS yields sequence information. Peptides are ionized as they enter the mass spectrometer and scanned for parent ion mass. In tandem quadrupole instruments, ions of interest are then selected for sequencing by electric field guidance into a collision cell, where ions are activated by collision with an inert gas; CID and fragmentation result. In ion trap instruments, ions of interest are concentrated by field guidance into an electromagnetic trap, where they are accelerated in the presence of helium, eventually accumulating sufficient energy through multiple collisions to allow peptide fragmentation. From the fragmentation patterns, sequence information can be derived and enable identification of the peptide sequence and represented protein.

highly accessible technique. Trypsin is commonly chosen for proteolysis because it is efficient and predictable and produces many peptides optimal for recovery from gels and MS/MS fragmentation. In-gel digests were initially analyzed by direct infusion at low flow rate with electrospray ionization (ESI), in which analytes are introduced to the MS inlet in liquid droplets sprayed from a needle in a high-voltage field (*7, 8*). Alternative methods combined proteolytic digests with a matrix, spotting many samples onto a plate that is then inserted into a mass spectrometer interface. A laser is used to volatilize the matrix, carrying the peptides into the gas phase for MS analysis; thus, this method is referred to as matrix-assisted laser desorption ionization (MALDI) (*9*).

2D gels are attractive because they provide direct information on protein abundances and modifications, and several outstanding studies have used 2DE to identify novel cancer biomarkers or targets of signaling pathways (*10–13*). Approaches that combine protein enrichment by fractionation with narrow-range p*I* separations or large-format 2D gels (*14, 15*) allow separation of up to ~10,000 features ("spots") that can report sequence variants and modified forms of proteins. However, the method is limited in overall sensitivity as well as detection of hydrophobic, basic, or large proteins, and so far no study has identified >1000 different gene products from a single cell type or tissue (*e.g.*, http://proteomics.cancer.dk).

An alternative to 2DE involves separation of proteins into fractions (*e.g.*, multidimensional chromatography, subcellular fractionation, or 1DE), followed by solution proteolysis to generate peptides. The outlet of a reversed-phase (RP) high-performance (or high-pressure) liquid chromatography (HPLC) column can be directly coupled to the ESI inlet of the mass spectrometer to analyze complex peptide mixtures more efficiently, a method referred to as LC/MS or LC/MS/MS (*16, 17*). This is often referred to as microionspray or nanospray ionization, when adapted to low flow rate. RP chromatography is ideal, because ESI requires that the samples be delivered in a volatile solvent. Large numbers of peptides can be sequenced in a single run by coupling RP columns to a mass spectrometer capable of automated MS/MS triggered by MS peak detection (*18, 19*).

A limitation of LC/MS/MS of complex samples is that peptides elute more quickly than currently available mass spectrometers can collect MS/MS spectra. This can be addressed by chromatographic prefractionation of peptides prior to RP. The most common methods utilize ion exchange or isoelectric focusing, either by coupling columns in tandem with RP (*20–24*) or by collecting fractions off-line. Another approach is to analyze the sample several times by using different, overlapping narrow mass windows ("gas-phase fractionation", *25*). Early applications of multidimensional LC/MS/MS to complex biological samples reported high-throughput sequencing and identification of proteins in ribosomal complexes and cell lysates from yeast (*23, 24*).

Because these strategies involve digestion of complex samples containing many proteins, they are often referred to as "bottom-up" or "shotgun" methods. In contrast, 2DE represents a "top-down" approach, because individual proteins are separated and compared before MS analysis. One related top-down strategy introduces intact proteins by ESI or MALDI to the mass spectrometer and then profiles differences between samples. This is used in surface enhanced laser desorption ionization (SELDI, *26*), where protein mixtures are simplified by affinity enrichment on sample plates and then ionized by MALDI. A major limitation of top-down methods is that inefficient fragmentation of large analytes by CID often precludes protein identification, and most studies so far have been restricted to high p*I* and/or low-mass proteins (*27, 28*). However, recent electron capture dissociation (ECD) or electron transfer dissociation (ETD) methods, when combined with high-resolution MS (FTICR MS) show promise in alleviating this problem (*29–31*). Fragmentation involves electron capture by the analyte followed by free-radical-based backbone cleavage, which is largely sequence-independent. A recent top-down study reported 101 protein identifications from bacterial whole-cell lysates (*32*), and an analysis of histones illus-

trates the power of this approach for revealing complex modifications of proteins and combinatorial regulation of modification sites (*33*).

In-Depth Protein Identification by Bottom-Up Proteomics. How feasible is it to identify every protein in highly complex samples, such as whole-cell extracts? In-depth profiling of all proteins is most efficiently accomplished by solution digestion, because every protein contains at least some peptides that are easily sequenced by MS/MS. A limiting factor in detecting peptides is the need to isolate each peptide in the mass spectrometer for fragmentation. The number of different expressed gene products varies with cell type, but in mammalian cells they can reasonably be estimated at ~12,000 on the basis of studies of mRNA complexity in mouse tissues and massively parallel signature sequencing in human cell lines (*34, 35*). This predicts ~420,000 peptides generated by tryptic digestion that are within the mass detection range of commonly used mass spectrometers and chromatographically resolvable with commonly used HPLC systems. On average, peptides are observed in 2 charge states, and split distributions of peptides or proteins between adjacent fractions results in ~4-fold resampling. Thus, comprehensive sampling would require resolving ~3.4 million ions, even without considering the detection of additional peptide forms due to splice variants and covalent modifications.

Although the high resolving power and scanning speeds of current MS instruments suggest that at least one peptide per protein could be observed in complex mixtures with relatively little separation, in practice, peptide identifications fall short of surveying all proteins. Recent studies examining the composition of mammalian cells typically report identifications of ~2000 proteins, with some examples reaching up to ~5000 proteins (*36–40*). One complication is that useful peptide identifications are generated from only ~10% of MS/MS attempts. This occurs in part because of uneven distribution of peptides across the sampling space and because MS/MS spectra are often collected on noise peaks or fragment ions generated outside the ion trap/collision cell, which "distract" the instrument from sequencing true parent ions. In addition, sample complexity causes important limitations for in-depth profiling. Weak ions that coelute with intense ions may be missed when their intensity differences exceed the instrument dynamic range. MS/MS isolation windows of ~2 Da are typically used to achieve a high signal to noise ratio, which leads to MS/MS of two or more coeluting ions with the same m/z; complex "chimera" spectra are generated that are usually difficult to assign. Therefore, the ability to separate ions (peak capacity), the rate of data collection (sampling rate), and the instrument sensitivity are important technical limitations to address.

Of these problems, those due to ion coelution can be minimized by sample fractionation. To the extent that peptide mixtures can be simplified to separate elution of weak *versus* intense ions, the sample loading can be scaled to observe peptides present at low copy number. Dynamic fill options, which limit the number of ions in the trap/collision cell, can be used to limit saturation at high loading. However, as more fractions are required to sample lower-abundance peptides, in-depth profiling becomes very time-consuming. Solutions are needed to improve peak capacity by increasing analyte resolution. Long RP columns (*e.g.*, 80 cm) run at ultra-high pressure and have been reported to yield ~7-fold greater peak capacity (*41*), and ultra-high-pressure chromatography systems should improve sampling as they become robust enough for nonstop operation. Also under development are ion mobility spectrometry (IMS) methods, where analytes are separated in the gas phase before fragmentation (*42, 43*), which when commercialized may have a significant impact toward solving peak capacity limitations.

Computational Algorithms for Peptide Identification. Proteomics profiling relies on search programs that match MS/MS spectra to candidate peptides in an automated manner. Detailed descriptions of current algorithms can be found in several recent reviews (*44–47*). The most successful algorithms calculate a score of some type that evaluates similarity between the observed

KEYWORDS

Collision-induced dissociation (CID): A mechanism for fragmenting ions in the gas phase by collisions with a neutral species. Collisions convert ion kinetic energy to internal energy, and this leads to the formation of fragment ions.

MS/MS spectrum: This spectrum of fragment ion masses is a two-stage process of isolating a precursor by mass followed by CID or other fragmentation methods and analysis of the fragments.

MS^3 spectrum: An extension of MS/MS, in which one or more of the fragment ions are selected and further cleaved by CID or other fragmentation methods. The resulting MS^3 fragment ions reflect the composition of the MS/MS fragment ions. MS^3 is often helpful when MS/MS cleavage is inefficient.

Electrospray ionization (ESI): A method that ionizes analytes and efficiently transfers nonvolatile biomolecules from solution to the gas phase. Analytes are introduced to the MS inlet at atmospheric pressure *via* liquid droplets sprayed from a finely drawn needle held at a high potential relative to the inlet. Large electric fields generate charged droplets. Evaporation of solvent from the droplets leads to increased charge density and coulombic repulsion between analytes, driving fission processes that eventually create single, gas-phase analyte molecules.

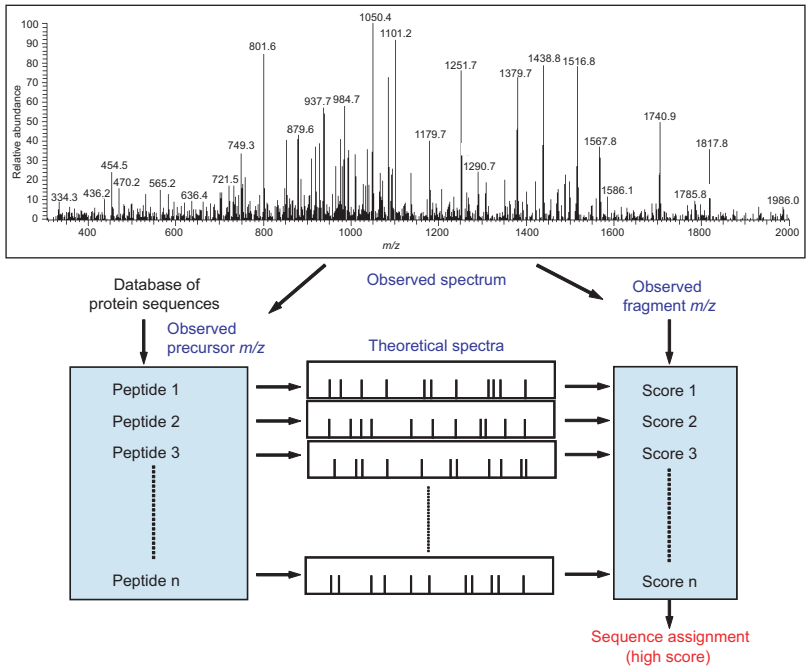

Figure 2. Similarity scoring by search programs. For each MS/MS spectrum, a list of peptide candidates is selected on the basis of parent ion mass. Theoretical spectra corresponding to each sequence are then scored for similarity with the observed MS/MS spectrum. Peptides are ranked according to score, with the top scoring sequence chosen as the most likely assignment. Currently, most programs match fragment masses but not intensities when comparing observed *vs* theoretical spectra because of difficulties in quantitative prediction of relative intensities. However, new studies reveal improved discrimination between correct and incorrect assignments when intensity information can be predicted and assessed.

MS/MS spectrum and a theoretical spectrum (*e.g.*, Sequest) or that uses a statistical approach to evaluate the probability of observing the MS/MS fragment ions (*e.g.*, Mascot) (Figure 2). A recent study reported a measurable advantage of the probability-based Mascot program over four others (Sequest, Spectrum Mill, Sonar, and X!Tandem) with respect to the data capture achieved with constant false positives (*48*). This agrees with our own experience, where highest data capture was observed with OMSSA (*49*), followed by Mascot, then Sequest and X!Tandem (Resing *et al.*, unpublished results).

The most widely used search programs often show low discrimination between true and false assignments. Researchers may be tempted into lowering score thresholds in order to capture more data, but this leads to acceptance of many false assignments. The problem is aggravated in the search for post-translational modifications or nonspecific proteolytic products, which increases the effective search space. As a result, studies may show large variability in the percentages of assignments that are false but accepted and large numbers of assignments that are true but rejected. A common practice is to manually evaluate assignments with borderline scores, although care must be taken to avoid biasing results based on assumptions that certain proteins are present in the sample.

Orthogonal metrics are useful for filtering incorrect assignments. For example, Peptide Prophet uses a linear discriminant strategy to develop weighting coefficients for different scores generated by Sequest in order to improve confidence of matches (*50*). Using the difference between scores of the top two highest ranked sequences as a filter also allows score thresholds to be lowered without increasing the false-positive rate (*51*). Voting methods that evaluate consensus in sequence assignments between two search programs also improve accuracy, because they allow data capture at reduced score thresholds (*40*). In addition, evaluation of peptide chemical properties can minimize inclusion of unlikely peptide assignments. For example, effective ways to remove false assignments are to evaluate the likelihood of observing incomplete proteolysis products or the concordance between sequence and chromatographic behavior for candidate peptides (*40, 52*).

Most algorithms for similarity scoring evaluate fragment ion masses without considering relative fragment ion intensities. These would benefit from improved methods to predict theoretical spectra, because fragment ion intensities reflect rates of cleavage at different bonds and therefore contain significant information about the chemical plausibility of a candidate sequence. To address this, one approach analyzes fragment ion intensities in MS/MS spectral databases, either by statistically evaluating intensities representing bond cleavages between specific pairwise combinations of amino acids (*53*) or by using machine learning methods to model intensity patterns (*54*). Another approach utilizes a library of observed MS/MS spectra that have been averaged to remove noise and chemical contamination (*55, 56*). Similarity scoring against spectral libraries shows good results, although the libraries so far contain only ~6% of possible peptide sequences (3–4% of residues in yeast and human proteins). This is because often only a subset of tryptic peptides from

each protein shows highly efficient fragmentation and is thus more highly represented in MS/MS searches, with lower coverage expected for proteins that are less extensively characterized. Thus, from an operational point of view, the success of the libraries depends on repeated observation of these subsets in different experiments.

A very promising approach is the prediction of theoretical spectra based on chemical models of gas-phase fragmentation, including the prevalent "mobile proton" mechanism involving protonation of a backbone carbonyl oxygen followed by subsequent cleavage of the adjacent peptide bond (*57–59*). Zhang (*60, 61*) developed a kinetic model for known cleavage events and then fit parameters describing these processes with varying amino acids and gas-phase basicities by using a large training set of MS/MS spectra. The optimized parameter set was used to simulate MS/MS fragmentation from peptide sequences. Good agreement was reported between theoretical and observed spectra, and our lab has corroborated the performance of this algorithm in predicting spectra in shotgun datasets (*62*). Although chemistries that deviate from these models are to be expected, the results indicate that current models of gas-phase fragmentation chemistries are mature enough to quantitatively predict relative fragment ion intensities in most cases, yielding improved discrimination of peptide assignments.

The Problem of Protein Inference. Ambiguity ensues with protein "isoforms", which share peptide sequences that cannot be distinguished. These include identical sequences, sequences with isobaric changes in residue (replacing isoleucine for leucine), or those with mass differences of 1 Da where MS/MS is collected on the second isotope peak (which may resemble replacement of aspartic acid for asparagine or glutamic acid for glutamine). Often when the same peptide sequence or a close isoform is found in more than one protein, replicate MS/MS spectra representing the same peptide sequence are assigned different protein accession numbers. The same peptide sequence in a different sample may be matched to a different protein entry; this can introduce error when samples are compared and lead to inflated protein counts.

Four algorithms exemplify different approaches to evaluate whether a peptide assignment is found in more than one protein entry. DTASelect and ProteinProphet use a data-dependent method for resolving those ambiguities, considering the protein assignments made and looking for isoform "overlaps" (*63, 64*). IsoformResolver and DBParser invert the process by creating peptide-centric databases, where each distinct peptide sequence is represented once and then linked to all protein entries containing that sequence (*40, 65*). In this approach, only the peptide sequence is taken from the search result, and protein assignments are made by processing sequences against the peptide database. With IsoformResolver, ~24% of protein entries from a profile of a human sample could be removed by eliminating redundancies between isoforms, and an additional 5% of proteins could be removed by eliminating unlikely peptide isoforms with Ile/Leu or Asp/Asn substitutions. The need to unambiguously determine the composition of different protein isoforms is an important concern when differences in protein abundances between samples are compared.

Quantifying Changes in Protein Abundance. *Stable Isotope Labeling.* Bottom-up proteomics requires ways to report changes in protein abundance from peptide measurements. A common strategy is to use stable isotope labeling (*e.g.*, ^{12}C *vs* ^{13}C, ^{14}N *vs* ^{15}N, ^{2}H *vs* ^{1}H) and then mix the samples to allow internal comparison of peptides between samples (Figure 3). Relative changes in protein abundance can then be determined from the ratio of intensities between differential isotopically labeled peptides. Proteins can be covalently labeled with stable-isotope-labeled moieties (*e.g.*, cysteine-alkylating reagents, lysine-coupling with methylisothiourea, and terminal carboxyl labeling with ^{16}O *vs* ^{18}O-labeled water) or metabolically labeled with isotopically distinguishable amino acids, allowing pairwise mixtures to be compared (*66–70*). Multiplexed iTRAQ (isobaric tags for relative and absolute quantitation) reagents have also been designed that allow chemical coupling of isobaric adducts onto peptides, which then yield isotopically distinguish-

KEYWORDS

Matrix-assisted laser desorption ionization (MALDI): A laser ionization method efficient for biomolecular analytes, particularly for more hydrophobic species. Analytes are co-crystallized with a matrix and deposited onto a plate. UV laser activation of the matrix carries some of the analytes, usually ionized with a single charge, into the gas phase for MS detection.

Surface-enhanced laser desorption/ionization (SELDI): Analytes are first simplified by differential enrichment on solid supports *via* affinity or ion exchange capture and washes of varying stringency, then ionized from the same plates by MALDI. SELDI is advantageous for its simplicity of analysis but is limited by the dynamic range of the MALDI method, the resolution of the mass spectrometer, and its inability to directly identify the analytes.

Quadrupole MS: A mass analyzer configuration in which ions travel between four parallel metal rods to which a radio frequency field is applied. By scanning voltages, users can select ions with varying m/z (typically 15–2000 Da) for transmission to a detector and monitoring. A tandem quadrupole configuration is often used to perform MS/MS.

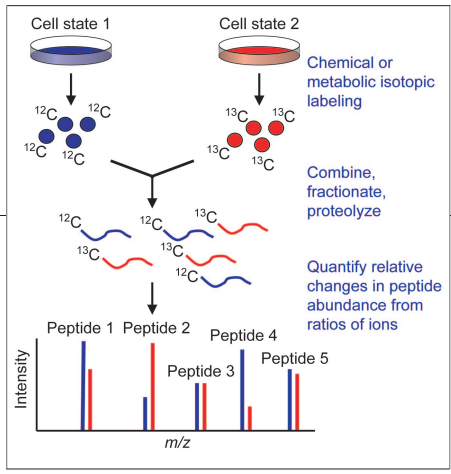

Figure 3. Quantifying relative changes in protein abundance by stable isotope labeling. Proteins from different samples are differentially tagged with isotopically labeled moieties, which include amino acids incorporated by metabolic labeling, or chemical adducts coupled covalently following cell disruption. Differentially labeled proteins are mixed and proteolyzed, and peptides are analyzed by MS. Relative changes in protein abundance are reflected by ratios of peak intensity for matched isotopically labeled peptides.

able fragment ions after MS/MS (*71*). Published studies have demonstrated measurements of protein abundance changes down to 30–50%, with high accuracy and reproducibility in simple mixtures. Chemical reactivity may vary depending on peptide abundance and thus may increase the variability of ratio measurements (*72*); as a result, metabolic labeling is gaining in popularity. Effects of metabolic labeling efficiency and other variables of this experiment on data interpretation have been comprehensively reviewed (*73*).

Quantitation of Peak Area Intensities. Label-free quantitation methods are promising alternative strategies, and recent studies demonstrate the feasibility of quantifying protein abundance ratios from peak area intensities, measured in different samples that are unmixed and run in parallel by LC/MS/MS. Although earlier assumptions held that variable ion suppression of peptides coeluting with different analytes would preclude accurate quantitation, recent evidence reveals the success of the approach when reproducibility of chromatography can be maintained. Stringent statistical methods and replicate analyses are key considerations when protein changes are identified. As with stable isotope labeling, the approach works less well for low-abundance proteins, because of the need to accurately define peak boundaries and signal to noise ratio. In addition, intensity assessment can be compromised when ions with similar mass coelute, which is often seen with complex mixtures. For these reasons, measurements are most reliable when the sequence coverage of a given protein is high enough to identify ≥3 peptides common to each sample. Initial studies demonstrated linearity of peptide intensities for protein standards spiked into serum or plasma (*74, 75*). Other reports have confirmed these results in shotgun datasets (*76–78*), and for peptides separated off-line by ion exchange HPLC followed by RP LC/MS/MS, protein ratios of ~2-fold or more can be quantified with high confidence after summing over multidimensional steps. Thus, although the approach is not as precise as stable isotope labeling, it nevertheless can be used to address many types of biological questions, including those for which labeling is not feasible.

LC/MS Quantitation Followed by Targeted MS/MS. An approach that we will refer to as differential feature detection (DFD) MS first surveys LC/MS peaks to profile ions showing differences in intensity and then reanalyzes samples by LC/MS/MS to selectively target those ions for sequencing (*79, 80*). Advantages are that the complexity of ions for MS/MS analysis can be reduced to those that are significantly altered between samples, analogous to differential protein profiling by 2DE. These methods are related to "accurate mass tag" concepts pioneered by Richard Smith, in which proteins are profiled on the basis of accurate measurement of peptide masses and peak intensities (*81*). Although many MS instruments enable simultaneous collection of MS and MS/MS data, the approach often requires a separate LC/MS/MS run to target specific ions that may not have been sampled initially. Computational algorithms that are able to accurately align and profile features between many separate runs are essential for this approach and represent an area of active development (*82–84*).

DFD-MS requires very reproducible chromatography and repeated runs to gain statistical confidence in peak intensity changes, given that peptides for each protein cannot be grouped beforehand by sequence. Thus, current applications involve separation of each sample by a single RP run, although peak capacity considerations indicate that high coverage of complex peptide mixtures will require prefractionation by multidimensional chromatography. DFD-MS has considerable potential for analyzing samples with lower complexity than whole-cell extracts, especially those for which labeling is difficult. Applications to biomarker discovery in human patient fluids is underway in pharmaceutical and academic groups, and initial results seem very promising.

Spectral Counting. This approach sums the total number of MS/MS signals of any peptide in a given protein, observed as different sequencing attempts, charge states, or elution in different chromatographic fractions. It is the simplest method to apply in experiments requiring multidimensional peptide separations, and it assumes random sampling statistics. Protein ratios have been calculated from spectral count information by adapting mathematical expressions used in

serial analysis of gene expression that correct for differences in total spectra between samples and eliminate discontinuity in ratio estimates (*76*). With these methods, protein ratios 2-fold or greater could be estimated from shotgun datasets, where the presence of ≥4 spectra/protein in at least one of the samples was required to estimate ratios with high confidence.

Ratios measured by spectral counts show higher dynamic range than direct peak intensity or stable isotope ratio measurements (*76, 85, 86*). Thus, spectral counts more accurately quantify large changes in abundance under conditions where incomplete sampling can be assumed, whereas peak area intensity measurements yield better estimates of smaller changes, with greatest accuracy occurring as sampling approaches saturation. For these reasons, the strengths of spectral counting complement other quantitation methods.

A study comparing four replicate analyses of yeast proteins labeled in rich (^{14}N-labeled) *versus* minimal (^{15}N-labeled) media showed greater reproducibility between replicates when measuring ratios from spectral counts than ratios of ^{14}N/^{15}N ions averaged over many peptides (*86*). This was attributed to large variations in ^{14}N/^{15}N ion chromatogram ratios between different peptides corresponding to a given protein. Thus, global measurements by spectral counting appeared more reproducible than the isotope labeling measurements that, while more precise, must be averaged over individual ions. It will be a surprising but welcome trend if the simplest and most accessible method for quantitation turns out to perform best for many applications.

Multiple-Reaction Monitoring. An important goal for proteomics is to quantify protein abundances in many different samples (*e.g.*, clinical specimens) to enable statistical evaluation across populations. Large numbers of samples cannot easily be surveyed by global profiling, which requires the use of a candidate-based approach in which a restricted set of targets are selected for quantitation. Multiple-reaction monitoring (MRM) MS is a multiplexed form of selected ion monitoring that selectively observes an analyte by using two mass filters, first scanning the mass of the parent ion and then transitioning to scanning one or more fragment ions generated by MS/MS (*87–89*). Analytes can be quantified from peak heights of fragment ions, and this has the advantages of achieving greater sensitivity over full-scanning methods because of the restriction in mass range and avoiding interference from coeluting molecules in complex mixtures. When isotopically labeled internal standards are added, changes in absolute concentrations can be obtained from fragment ion current.

Although MRM has long been used for detection of small molecules such as metabolites or drugs, interest is growing in quantifying proteins based on peptide MRMs, for example, in investigating polypeptide markers in patient fluids. In this approach, peptide and fragment ion masses that uniquely identify specific proteins are calculated, and experimental MS and MS/MS spectra are used to identify subsets of peptides and fragment ions with best intensity, which are then selectively monitored. A recent analysis performed 137 MRMs on high- and medium-abundance proteins in human plasma, and 47 proteins were quantified between replicate runs (*89*). The results showed proteins that were detectable to ~1 μg/mL, with high precision over a dynamic range of ~5×10^4.

Recent Applications. Among the best proteomics studies are those that focus on achieving comprehensive coverage, seek to quantify temporal or comparative differences, are statistically rigorous, apply conservative rules to data analysis, and validate conclusions by independent means. Methods developed to date have shown the greatest success in defining the protein compositions of samples that can be simplified by purification or enrichment. We describe two successful applications of proteomics technologies, profiling proteins within subcellular organelles and cellular phosphoproteins.

Organelle Composition. An ambitious global proteomics study carried out comparative profiling of four subcellular compartments from six mouse organs (*37*). More than 8 million spectra from 203

KEYWORDS

Ion trap MS: Ions are collected in a 3D or 2D electrostatic trap to which a radio frequency field is applied. Voltages can be varied to focus and concentrate specific ions in the trap while all others are ejected. CID can be achieved by applying an excitation voltage to the trap endcaps at a frequency in resonance with the oscillation frequency of a precursor of interest.

Time-of-flight MS (TOFMS): A mass analyzer configuration in which ions are accelerated through a large electric field and then allowed to travel through a field-free flight tube, monitoring the time taken to reach a detector. Because velocity of travel is inversely proportional to mass$^{1/2}$, lighter ions reach the detector more quickly than heavier ions.

Fourier transform MS (FT-MS): An MS configuration that detects signals on metal plates from electrical currents produced by ion oscillations. The hallmark of FT-MS is its high resolution; 1–2 ppm mass accuracy or lower can be achieved. A drawback is the long cycle time required for acquisition of MS and particularly MS/MS spectra.

Electron capture dissociation (ECD) and electron transfer dissociation (ETD): These mechanisms are used for gas-phase fragmentation involving electron capture by an analyte to form a free radical with high energy. Peptides primarily undergo cleavage at backbone N-αC bonds to generate N-terminal "c" ions and C-terminal "z" ions. ETD interfaced to an ion trap eliminates the cutoff problem of CID fragment ions and allows for the observation of low-mass ions.

Validation remains a rate limiting step compared with the wealth of information revealed by proteomics.

LC/MS/MS experiments were used to define the subcellular localizations of >3000 proteins, half of which were previously uncharacterized. To accomplish this, the researchers subjected the data to rigorous statistical filtering and machine learning analyses and validated localizations of a handful of the new assignments by independent means. The study illustrated how relative abundances of different proteins could be determined by spectral counts where 70% of cases were consistent with mRNA expression measurements. The compendium of data provides a rich source for mining biologically meaningful information for both organelle and tissue expression of proteins.

Comprehensive profiling of specific subcellular compartments were the goal of several investigations that have yielded insights into the dynamic nature of organellar protein composition. An analysis of the yeast outer mitochondrial membrane accounted for 85% of known outer mitochondrial proteins (*90*). Of note in this study is the improved sampling obtained by LC/MS/MS (85% of known yeast outer membrane proteins) compared with a 2DE strategy optimized for membrane proteins (31%). This study also defined a subset of "preproteins" in the outer mitochondrial membrane, destined for inner mitochondrial compartments.

Such analyses will invariably raise doubts about proteins that appear to be mislocalized. One strategy for circumventing this problem is to use MS to profile organelles separated by traditional fractionation techniques, such as sucrose gradient centrifugation, rather than purified organelles. Global approaches to organellar profiling start with crude preparations (*91, 92*), relying on differential density centrifugation of subcellular components and cosedimentation of organelle marker proteins to classify localization. In this way, the *Arabidopsis* organellar proteome was mapped with an iTRAQ stable-isotope-labeling method (*91*), and mouse liver organellar proteins were mapped with label-free peptide ion intensity measurements (*92*). Subcellular localization was confirmed by independent means for only a few of the hundreds of proteins identified in these two studies; this illustrates a general problem that validation remains a rate limiting step compared with the wealth of information revealed by proteomics.

A series of elegant studies examined the composition of human nucleoli and quantified the dynamics of exchange in response to metabolic inhibitors (*93, 94*). Nucleoli were purified from HeLa cells, and proteins were separated into 16–20 fractions by 1DE. After in-gel digestion and LC/MS/MS, unique peptides were identified, and high-accuracy mass determinations were used to help confirm the assignments. The study reported 692 proteins by 11,130 unique peptides, and they were reproducibly observed between different organelle preparations. These overlapped with 87% of yeast proteins previously identified in nucleoli by global tagging strategies (*95*), an indication that the proteomics sampling level was close to saturation. Many proteins were novel or previously uncharacterized and were further examined as chimeras tagged with YFP. Of 18 YFP fusion proteins tested, 15 were localized to nucleoli, a validation of the ability of the proteomics screen to identify novel components of this organelle (*93*).

The dynamics of nucleolar proteins were then profiled with stable isotope labeling with amino acids in cell culture (SILAC) with triple-coded isotopomers, and cells arrested in transcription were monitored with actinomycin D for varying times (*94*). Of the 489 proteins that could be quantified, about one-third showed significant changes in abundance within 150 min, which could be ascribed to redistribution in and out of nucleoli. Interestingly, the profiles showed coordinate kinetics of proteins with shared functions, an indication that nucleoli do not exchange components randomly but rather transition between definable states. Most likely, this reflects unknown mechanisms for controlling removal or recruitment of functional complexes in response to cell stress.

Phosphoproteomics. The comprehensive annotation of the phosphoproteome is entering a period of explosive growth because of recent advances in MS instrumentation and methods. Recently, it has become common for a single study to identify hundreds or thousands of phosphorylation sites, many of which were previously uncharacterized (*96–102*). Phosphopeptides

KEYWORDS

Ion mobility spectrometry (IMS): Ions pass through a drift tube filled with a neutral buffer gas and charged with a uniform electric field. The ions resolve on the basis of their mobility as a function of analyte mass, size, and shape. Tandem IMS-MS configurations enable separation of precursor ions, followed by precursor ion selection and CID. Advantages are that complex mixtures of analytes can be separated prior to the mass spectrometer by a method orthogonal to reversed-phase HPLC.

Mobile proton mechanism: A model to describe a common chemical mechanism for gas-phase peptide fragmentation in positive ion MS. Peptides acquire charge by protonation in an ion source, and protons rapidly redistribute among different atoms in the molecule according to relative basicity. Protonation of the nth carbonyl oxygen on the peptide backbone enhances the rate of attack by the n−1 carbonyl oxygen to form a cyclized oxazolone. Subsequent CO–NH bond cleavage yields a "b" ion containing the N-terminus, and a "y" ion containing the C-terminus.

are difficult to identify in complex peptide mixtures for several reasons. Phosphoserine and phosphothreonine are labile upon peptide ionization; this alters the chemistry of peptide fragmentation, with much of the ionizing energy frequently absorbed in the neutral loss of H_3PO_4 rather than peptide bond fragmentation (*103*). Neutral loss of HPO_3 also sometimes occurs with phosphotyrosine, phosphoserine, and phosphothreonine (*103*). This complicates the problem of correctly identifying the peptide as well as determining occupied sites on peptides with more than one phosphorylatable residue. An emerging approach to alleviate this problem is the use of ETD, which allows rapid fragmentation and retention of post-translational modifications on the fragment ions (*104*). Computational methods have yet to satisfactorily address various obstacles in phosphopeptide identification, which relies heavily on manual validation. Thus, current limitations in data analysis present a barrier to large-scale phosphorylation site mapping.

Low-phosphorylation stoichiometry, especially on low-abundance proteins, increases the difficulty of sampling phosphopeptides from complex mixtures. Approaches to solve this problem exploit the unique chemical properties of phosphate groups in order to enrich phosphopeptides relative to high background unphosphorylated peptides. Most studies combine separation techniques with biochemical methods for affinity selection. For example, many studies have used immobilized metal ion affinity chromatography (IMAC) or phosphospecific antibody-coupled resins to enrich phosphopeptides or phosphoproteins, followed by RP LC/MS/MS (*97–102*). Other enrichment strategies capitalize on the unique chemical reactivity of phosphoester bonds on phosphopeptides, often exploiting the β-elimination reaction of phosphoserine or phosphothreonine, which allows subsequent addition of tagged adducts at the site of phosphorylation (*105, 106*). An alternative chemistry involves the reversible capture of phosphopeptides by amine-containing solution polymers (*107*). However, chemical coupling methods often involve complex sample handling and have not yet proven successful for in-depth profiling.

Quantitation of phosphorylation dynamics in a biologically meaningful timescale is a major area for development. Stable isotope labeling is often used to quantify changes in phosphopeptides and is currently the best means for circumventing problems due to irreproducible recovery from affinity columns (*108–110*). However, the approach can be difficult to reproduce, as illustrated by a study of tumor necrosis factor-α signaling in HeLa cells, where proteins from stimulated and unstimulated cells were reciprocally labeled with ^{14}N and ^{15}N stable isotopes followed by IMAC enrichment of phosphopeptides (*111*). Although many phosphopeptides were observed in common between the two experiments, none of the phosphopeptides altered by ≥2-fold in one experiment could be quantified in the reciprocal experiment. Perhaps the biggest challenge with quantitation in phosphoproteomics is the necessary reliance on data from single peptides, which have lower statistical power compared with experiments in which multiple peptides can be quantified for the same protein. High-throughput experiments also require protein profiling to be carried out in parallel, in order to confirm that changes occur in phosphorylation stoichiometry rather than protein abundance.

The unique gas-phase fragmentation behavior of phosphopeptides upon CID can also be exploited to preferentially select for phosphopeptides, in essence, adding an enrichment step within the mass spectrometer. Some investigators set instruments so that only peptides demonstrating neutral loss of H_3PO_4 (−98 Da) will be selected for sequencing. Fragmentation of the neutral loss fragment generates an MS^3 spectrum that selectively represents the phosphopeptide, often reducing background from coeluting unphosphorylated peptides. MS^3 sequencing has been used to identify hundreds of phosphopeptides from mammalian cell systems (*96, 100*), although phosphorylated residues that do not exhibit neutral loss of H_3PO_4 (phosphotyrosine and about one-third of phosphoserine or phosphothreonine sites) will be unrepresented in the experiment.

A promising alternative strategy is to use negative ion mode detection to exploit the high intensity of phosphopeptides and derivatives. In one form of negative ion precursor scanning, the loss of negatively charged PO_3^- or $H_2PO_4^-$ ions from phosphopeptides yields relatively selective −79 Da and −98 Da signatures that can be traced back to the parent (*112, 113*). Switching from negative to positive ion mode then allows positive ion CID and sequencing of the phosphopeptide (Figure 4). Because unphosphorylated peptides are invisible in this experiment, phosphopeptides are preferentially identified and can be relatively quantified from the intensity of the −79 Da precursor ion. The approach shows signifi-

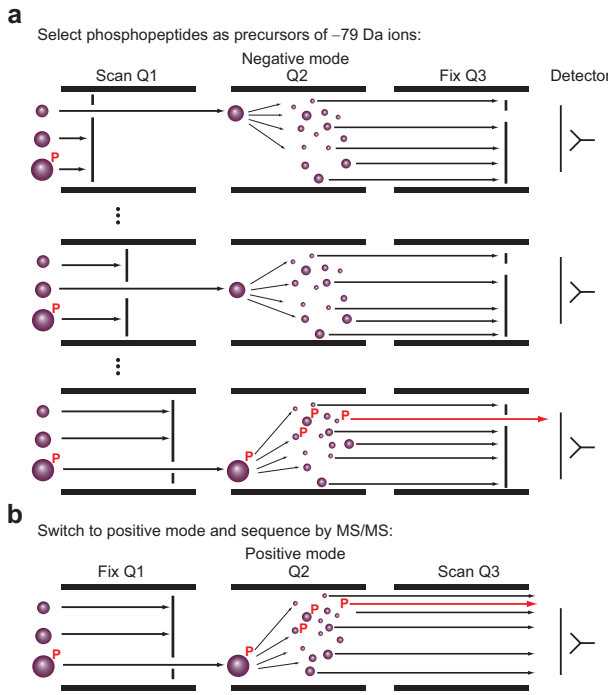

Figure 4. Identifying phosphopeptides by negative ion precursor scanning. This figure illustrates a method for manipulating ions to selectively sequence phosphopeptides in a tandem quadrupole mass spectrometer. a) Peptides and phosphopeptides eluting from nanoflow RP-HPLC are scanned for loss of PO_3^- in negative mode by scanning Q1, fragmenting in Q2, and monitoring for the phosphate marker ion (red "P") by fixing Q3 at −79 Da. b) When a phosphopeptide signal is detected, the polarity is switched to positive mode, the ion is selected in Q1 and fragmented in Q2. The ions are scanned in Q3 in linear trap mode to acquire the MS/MS spectrum for sequence identification. (Described in ref 112.)

cantly higher sensitivity for phosphorylation site identification compared with positive ion detection. However, it has not yet been widely applied to analysis of phosphopeptides in complex mixtures because of problems with ion suppression. Nevertheless, these and other methods illustrate how gas-phase manipulation of specific peptide chemistries may provide strategies for future development in profiling protein covalent modifications.

Summary. Proteomics is currently far from routinely achieving the depth of profiling enabled by DNA hybridization technologies. Nevertheless, these methods are needed to complement global mRNA measurements and to enable the analysis of samples that are inaccessible to technologies based on nucleic acids. Rapid progress is being made in solving the many technical limitations, such as incomplete sampling, peak capacity, and automated computational spectral analysis. The need to simplify complex mixtures and increase resolution of peptide and protein analytes is being addressed through improved methods for chromatography and biochemical separations, as well as advances in MS instrumentation that are generating faster scanning and more sensitive instruments.

An important emerging trend is the development of increasingly sophisticated methods to manipulate ions in the gas phase, within the mass spectrometers. For example, early proteomics experiments required proteins to be purified; this was superseded when automated triggering of MS/MS was made possible, followed later by MS^3, which allowed analysis of complex mixtures. The example of negative ion precursor scanning illustrates the utility of gas-phase manipulation for chemical selection. Future capabilities for IMS analyses of large peptides may also allow greater transfer of separation steps to the instruments. We expect such capabilities to eventually allow significant overlap between the mass range of analytes observable by top-down *versus* bottom-up strategies. As the analysis of larger polypeptides becomes feasible, the problem of sample complexity should be greatly simplified. Eventually, the goal of rapid protein profiling on a global scale will be possible.

Acknowledgments: We thank Karen Jonscher, University of Colorado Health Sciences Center, for helpful discussions.

REFERENCES

1. Celis, J. E., Gromov, P., Ostergaard, M., Madsen, P., Honore, B, Dejgaard, K., Olsen, E., Vorum, H., Kristensen, D. B., Gromova, I., Haunso, A., Van Damme, J., Puype, M., Vandekerckhove, J., and Rasmussen, H. H. (1996) Human 2-D PAGE databases for proteome analysis in health and disease: http://biobase.dk/cgi-bin/celis, *FEBS Lett. 398*, 129–134.
2. Domon, B., and Aebersold, R. (2006) Mass spectrometry and protein analysis, *Science 312*, 212–217.
3. Henzel, W. J., Watanabe, C., and Stults, J. T. (2003) Protein identification: the origins of peptide mass fingerprinting, *J. Am. Soc. Mass Spectrom. 14*, 931–942.
4. Mortz, E., O'Connor, P. B., Roepstorff, P., Kelleher, N. L., Wood, T. D., McLafferty, F. W., and Mann, M. (1996) Sequence tag identification of intact proteins by matching tanden mass spectral data against sequence data bases, *Proc. Natl. Acad. Sci. U.S.A. 93*, 8264–8267.

5. Zhang, W., and Chait, B. T. (2000) ProFound: an expert system for protein identification using mass spectrometric peptide mapping information, *Anal. Chem. 72*, 2482–2489.
6. Pappin, D. J. C., Hojrup, P., and Bleasby, A. J. (1993) Rapid identification of proteins by peptide mass fingerprinting, *Curr. Biol. 3*, 327–332.
7. Covey, T. R., Bonner, R. F., Shushan, B. I., and Henion, J. (1988) The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry, *Rapid Comm. Mass Spectrom. 2*, 249–256.
8. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, *Anal. Chem. 68*, 850–858.
9. Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991) Matrix-assisted laser desorption ionization of biopolymers, *Anal. Chem. 63*, 1193A–1203A.
10. Celis, J. E., and Gromov, P. (2003) Proteomics in translational cancer research: toward an integrated approach, *Cancer Cell 3*, 9–15.
11. Chen, G., Gharib, T. G., Wang, H., Huang, C. C., Kuick, R., Thomas, D. G., Shedden, K. A., Misek, D. E., Taylor, J. M., Giordano, T. J., Kardia, S. L., Iannettoni, M. D., Yee, J., Hogg, P. J., Orringer, M. B., Hanash, S. M., and Beer, D. G. (2003) Protein profiles associated with survival in lung adenocarcinoma, *Proc. Natl. Acad. Sci. U.S.A. 100*, 13537–13542.
12. Lewis, T. S., Hunt, J. B., Aveline, L. D., Jonscher, K. R., Louie, D. F., Yeh, J. M., Nahreini, T. S., Resing, K. A., and Ahn, N. G. (2000) Identification of novel MAP kinase pathway signaling targets by functional proteomics and mass spectrometry, *Mol. Cell 6*, 1343–1354.
13. Kanamoto, T., Hellman, U., Heldin, C. H., and Souchelnytskyi, S. (2002) Functional proteomics of transforming growth factor-beta1-stimulated Mv1Lu epithelial cells: Rad51 as a target of TGFbeta1-dependent regulation of DNA repair, *EMBO J. 21*, 1219–1230.
14. Zuo, X., and Speicher, D. W. (2002) Comprehensive analysis of complex proteomes using microscale solution isoelectrofocusing prior to narrow pH range two-dimensional electrophoresis, *Proteomics 2*, 58–68.
15. Klose, J., Nock, C., Herrmann, M., Stuhler, K., Marcus, K., Bluggel, M., Krause, E., Schalkwyk, L. C., Rastan, S., Brown, S. D., Bussow, K., Himmelbauer, H., and Lehrach, H. (2002) Genetic analysis of the mouse brain proteome, *Nat. Genet. 30*, 385–393.
16. Hunt, D. F., Henderson, R. A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E., and Engelhard, V. H. (1992) Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry, *Science 255*, 1261–1263.
17. Davis, M. T., Stahl, D. C., Hefta, S. A., and Lee, T. D. (1995) A microscale electrospray interface for on-line, capillary liquid chromatography/tandem mass spectrometry of complex peptide mixtures, *Anal. Chem. 67*, 4549–4556.
18. Courchesne, P. L., Jones, M. D., Robinson, J. H., Spahr, C. S., McCracken, S., Bentley, D. L., Luethy, R., and Patterson, S. D. (1998) Optimization of capillary chromatography ion trap-mass spectrometry for identification of gel-separated proteins, *Electrophoresis 19*, 956–967.
19. Figeys, D., and Aebersold, R. (1997) High sensitivity identification of proteins by electrospray ionization tandem mass spectrometry: initial comparison between an ion trap mass spectrometer and a triple quadrupole mass spectrometer, *Electrophoresis 18*, 360–368.
20. Anderegg, R. J., Wagner, D. S., Blackburn, R. K., Opiteck, G. J., and Jorgenson, J. W. (1997) A multidimensional approach to protein characterization, *J. Protein Chem. 16*, 523–526.
21. Opiteck, G. J., Lewis, K. C., Jorgenson, J. W., and Anderegg, R. J. (1997) Comprehensive on-line LC/LC/MS of proteins, *Anal. Chem. 69*, 1518–1524.
22. Figeys, D., Gygi, S. P., Zhang, Y., Watts, J., Gu, M., and Aebersold, R. (1998) Electrophoresis combined with novel mass spectrometry techniques: powerful tools for the analysis of proteins and proteomes, *Electrophoresis 19*, 1811–1818.
23. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R. (1999) Direct analysis of protein complexes using mass spectrometry, *Nat. Biotechnol. 17*, 676–682.
24. Washburn, M. P., Wolters, D., and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat. Biotechnol. 19*, 242–247.
25. Spahr, C. S., Davis, M. T., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Courchesne, P. L., Chen, K., Wahl, R. C., Yu, W., Luethy, R., and Patterson, S. D. (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest, *Proteomics 1*, 93–107.
26. Engwegen, J., Gast, M., Schellens, J., and Beijnen, J. (2006) Clinical proteomics: searching for better tumor markers with SELDI-TOF mass spectrometry, *Trends Pharmacol. Sci. 27*, 251–259.
27. Louie, D. F., Resing, K. A., Lewis, T. S., and Ahn, N. G. (1996) Mass spectrometric analysis of 40 S ribosomal proteins from Rat-1 fibroblasts, *J. Biol. Chem. 271*, 28189–28198.
28. Krishnamurthy, T., Davis, M. T., Stahl, D. C., and Lee, T. D. (1999) Liquid chromatography/microspray mass spectrometry for bacterial investigations, *Rapid Commun. Mass Spectrom. 13*, 39–49.
29. Cooper, H. J., Hakansson, K., and Marshall, A. G. (2005) The role of electron capture dissociation in biomolecular analysis, *Mass Spectrom. Rev. 24*, 201–222.
30. Lee, S. W., Berger, S. J., Martinovic, S., Pasa-Tolic, L., Anderson, G. A., Shen, Y., Zhao, R., and Smith, R. D. (2002) Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR, *Proc. Natl. Acad. Sci. U.S.A. 99*, 5942–5947.
31. Macek, B., Waanders, L. F., Olsen, J. V., and Mann, M. (2006) Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer, *Mol. Cell. Proteomics 5*, 949–958.
32. Patrie, S. M., Ferguson, J. T., Robinson, D. E., Whipple, D., Rother, M., Metcalf, W. W., and Kelleher, N. L. (2006) Top down mass spectrometry of <60-kDa proteins from Methanosarcina acetivorans using quadrupole FRMS with automated octopole collisionally activated dissociation, *Mol. Cell. Proteomics 5*, 14–25.
33. Beck, H. C., Nielsen, E. C., Matthiesen, R., Jensen, L. H., Sehested, M., Finn, P., Grauslund, M., Hansen, A. M., and Jensen, O. N. (2006) Quantitative proteomic analysis of post-translational modifications of human histones, *Mol. Cell. Proteomics 5*, 1314–1325.
34. Hastie, N. D., and Bishop, J. O. (1976) The expression of three abundance classes of messenger RNA in mouse tissues, *Cell 9*, 761–774.
35. Jongeneel, C. V., Iseli, C., Stevenson, B. J., Riggins, G. J., Lal, A., Mackay, A., Harris, R. A., O'Hare, M. J., Neville, A. M., Simpson, A. J., and Strausberg, R. L. (2003) Comprehensive sampling of gene expression in human cell lines with massively parallel signature sequencing, *Proc. Natl. Acad. Sci. U.S.A. 100*, 4702–4705.
36. Everley, P. A., Krijgsveld, J., Zetter, B. R., and Gygi, S. P. (2004) Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research, *Mol. Cell. Proteomics 3*, 729–735.
37. Kislinger, T., Cox, B., Kannan, A., Chung, C., Hu, P., Ignatchenko, A., Scott, M. S., Gramolini, A. O., Morris, Q., Hallett, M. T., Rossant, J., Hughes, T. R., Frey, B., and Emili, A. (2006) Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling, *Cell 125*, 173–186.

38. Patwardhan, A. J., Strittmatter, E. F., Camp, D. G., Smith, R. D., and Pallavicini, M. G. (2005) Comparison of normal and breast cancer cell lines using proteome, genome, and interactome data, *J. Proteome Res. 4*, 1952–1960.
39. Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., Yi, E. C., Dai, H., Thorsson, V., Eng, J., Goodlett, D., Berger, J. P., Gunter, B., Linseley, P. S., Stoughton, R. B., Aebersold, R., Collins, S. J., Hanlon, W. A., and Hood, L. E. (2004) Integrated genomic and proteomic analyses of gene expression in mammalian cells, *Mol. Cell Proteomics. 3*, 960–969.
40. Resing, K. A., Meyer-Arendt, K., Mendoza, A. M., Aveline-Wolf, L. D., Jonscher, J. R., Pierce, K. G., Old, W. M., Cheung, H. T., Russell, S., Wattawa, J. L., Goehle, G. R., Knight, R. D., and Ahn, N. G. (2004) Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics, *Anal. Chem. 76*, 3556–3568.
41. Shen, Y., Zhao, R., Belov, M. E., Conrads, T. P., Anderson, G. A., Tang, K., Pasa-Tolic, L., Veenstra, T. D., Lipton, M. S., Udseth, H. R., and Smith, R. D. (2001) Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics, *Anal. Chem. 73*, 1766–1775.
42. Liu, X., Plasencia, M., Ragg, S., Valentine, S. J., and Clemmer, D. E. (2004) Development of high throughput dispersive LC-ion mobility-TOFMS techniques for analysing the human plasma proteome, *Brief Funct. Genomic Proteomic 3*, 177–86.
43. Koeniger, S. L., Merenbloom, S. I., Valentine, S. J., Jarrold, M. F., Udseth, H. R., Smith, R. D., and Clemmer, D. E. (2006) An IMS-IMS analogue of MS-MS, *Anal. Chem. 78*, 4161–4174.
44. Sadygov, R. G., Cociorva, D., and Yates, J. R. (2004) Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book, *Nat. Methods 1*, 195–202.
45. Nesvizhskii, A. I., and Aebersold, R. (2005) Interpretation of shotgun proteomic data, *Mol. Cell. Proteomics 4*, 1419–1440.
46. Johnson, R. S., Davis, M. T., Taylor, J. A., and Patterson, S. D. (2005) Informatics for protein identification by mass spectrometry, *Methods 35*, 223–236.
47. Russell, S. A., Old, W., Resing, K. A., and Hunter, L. (2006) Proteomic informatics, *Int. Rev. Neurobiol. 61*, 127–157.
48. Kapp, E. A., Schutz, F., Connolly, L. M., Chakel, J. A., Meza, J. E., Miller, C. A., Fenyo, D., Eng, J. K., Adkins, J. N., Omenn, G. S., and Simpson, R. J. (2005) An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis, *Proteomics 5*, 3475–3490.
49. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H. (2004) Open mass spectrometry search algorithm, *J. Proteome Res. 3*, 958–964.
50. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search, *Anal. Chem. 74*, 5383–5392.
51. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, *J. Am. Soc. Mass Spectrom. 5*, 976–989.
52. 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., and Smith, R. D. (2006) Improved peptide elution time prediction for reversed-phase liquid chromatography-MS by incorporating peptide sequence information, *Anal. Chem. 78*, 5026–5039.
53. Huang, Y., Triscari, J. M., Tseng, G. C., Pasa-Tolic, L., Lipton, M. S., Smith, R. D., and Wysocki, V. H. (2005) Statistical characterization of the charge state and residue dependence of low-energy CID peptide dissociation patterns, *Anal. Chem. 77*, 5800–5813.
54. Gibbons, F. D., Elias, J. E., Gygi, S. P., and Roth, F. P. (2004) SILVER helps assign peptides to tandem mass spectra using intensity-based scoring, *J. Am. Soc. Mass Spectrom. 15*, 910–912.
55. Craig, R., Cortens, J. P., and Beavis, R. C. (2005) The use of proteotypic peptide libraries for protein identification, *Rapid Commun. Mass Spectrom. 19*, 1844–1850.
56. Frewen, B. E., Merrihew, G. E., Wu, C. C., Noble, W. S., and MacCoss, M. J. (2006) Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries, *Anal. Chem. 78*, 5678–5684.
57. Wysocki, V. H., Resing, K. A., Zhang, Q., and Cheng, G. (2004) Mass spectrometry of peptides and proteins, *Methods 35*, 211–222.
58. Wysocki, V. H., Tsaprailis, G., Smith, L. L., and Breci, L. A. (2000) Mobile and localized protons: a framework for understanding peptide dissociation, *J. Mass Spectrom. 35*, 1399–1406.
59. Paizs, B., and Suhai, S. (2005) Fragmentation pathways of protonated peptides, *Mass Spectrom. Rev. 24*, 508–548.
60. Zhang, Z. (2004) Prediction of low-energy collision-induced dissociation spectra of peptides, *Anal. Chem. 76*, 3908–3922.
61. Zhang, Z. (2005) Prediction of low-energy collision-induced dissociation spectra of peptides with three or more charges, *Anal. Chem. 77*, 6364–6373.
62. Sun, S., Meyer-Arendt, K., Eichelberger, B., Brown, R., Yen, C. Y., Old, W.M., Pierce, K., Cios, K., Ahn, N. G., and Resing, K. A. (2007) Improved validation of peptide MS/MS assignments using spectral intensity prediction, *Mol. Cell. Proteomics*, *6*, 1–17
63. Tabb, D. L., McDonald, W. H., and Yates, J. R. (2002) DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics, *J. Proteome Res. 1*, 21–26.
64. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry, *Anal. Chem. 75*, 4646–4658.
65. Yang, X., Dondeti, V., Dezube, R., Maynard, D. M., Geer, L. Y., Epstein, J., Chen, X., Markey, S. P., and Kowalak, J. A. (2004) DBParser: web-based software for shotgun proteomic data analyses, *J. Proteome Res. 3*, 1002–1008.
66. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat. Biotechnol. 17*, 994–999.
67. Brown, K. J., and Fenselau, C. (2004) Investigation of doxorubicin resistance in MCF-7 breast cancer cells using shotgun proteomics with proteolytic 18O labeling, *J. Proteome Res. 3*, 455–462.
68. Cagney, G., and Emili, A. (2002) De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging, *Nat. Biotechnol. 20*, 163–170.
69. Berger, S. J., Lee, S. W., Anderson, G. A., Pasa-Tolic, L., Tolic, N., Shen, Y., Zhao, R., and Smith, R. D. (2002) High-throughput global peptide proteomic analysis by combining stable isotope amino acid labeling and data-dependent multiplexed-MS/MS, *Anal. Chem. 74*, 4994–5000.
70. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, *Mol. Cell. Proteomics 1*, 376–386.
71. 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., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents, *Mol. Cell. Proteomics 3*, 1154–1169.
72. Parker, K. C., Patterson, D., Williamson, B., Marchese, J., Graber, A., He, F., Jacobson, A., Juhasz, P., and Martin, S. (2004) Depth of proteome issues: a yeast isotope-coded affinity tag reagent study, *Mol. Cell. Proteomics 3*, 625–659.

73. Beynon, R. J., and Pratt, J. M. (2005) Metabolic labeling of proteins for proteomics, *Mol. Cell. Proteomics 4*, 857–872.
74. Bondarenko, P. V., Chelius, D., and Shaler, T. A. (2002) Identification and relative quantification of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry, *Anal. Chem. 74*, 4741–4749.
75. Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T., Hill, L, Norton, S., Kumar, P, Anderle, M., and Becker, C. (2003) Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards, *Anal. Chem. 75*, 4818–4826.
76. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A., and Ahn, N. G. (2005) Comparison of label free methods for quantifying human proteins by shotgun proteomics, *Mol. Cell. Proteomics 4*, 1487–1502.
77. Wang, G., Wu, W. W., Zeng, W., Chou, C. L., and Shen, R. F. (2006) Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: reproducibility, linearity, and application with complex proteomes, *J. Proteome Res. 5*, 1214–1223.
78. Fang, R., Elias, D. A., Monroe, M. E., Shen, Y., 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., and Smith, R. D. (2006) Differential label-free quantitative proteomic analysis of Shewanella oneidensis cultured under aerobic and suboxic conditions by accurate mass and time tag approach, *Mol. Cell. Proteomics 5*, 714–725.
79. Wiener, M. C., Sachs, J. R., Deyanova, E. G., and Yates, N. A. (2004) Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures, *Anal. Chem. 76*, 6085–6096.
80. Bonneil, E., Tessier, S., Carrier, A., and Thibault, P. (2005) Multiplex multidimensional nanoLC-MS system for targeted proteomic analyses, *Electrophoresis 26*, 4575–4589.
81. Conrads, T. P., Anderson, G. A., Veenstra, T. D., Pasa-Tolic, L., and Smith, R. D. (2000) Utility of accurate mass tags for proteome-wide protein identification, *Anal. Chem. 72*, 3349–3354.
82. Ono, M., Shitashige, M., Honda, K., Isobe, T., Kuwabara, H., Matsuzuki, H., Hirohashi, S., and Yamada, T. (2006) Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry, *Mol. Cell. Proteomics 5*, 1338–1347.
83. Bellew, M., Coram, M., Fitzgibbon, M., Igra, M., Randolph, T., Wang, P., May, D., Eng, J., Fang, R., Lin, C., Chen, J., Goodlett, D., Whiteaker, J., Paulovich, A., and McIntosh, M. (2006) A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS, *Bioinformatics 22*, 1902–1909.
84. Jaffe, J. D., Mani, D. R., Leptos, K. C., Church, G. M., Gillette, M. A., and Carr, S. A. (2006) PEPPeR: a platform for experimental proteomic pattern recognition, *Mol. Cell. Proteomics 5*, 1927–1941.
85. Liu, H., Sadygov, R. G., and Yates, J. R. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics, *Anal. Chem. 76*, 4193–4201.
86. Zybailov, B., Coleman, M. K., Florens, L., and Washburn, M. P. (2005) Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling, *Anal. Chem. 77*, 6218–6224.
87. Covey, T., Maylin, G., and Henion, J. (1985) Quantitative secondary ion monitoring gas chromatography/mass spectrometry of diethylstilbestrol in bovine liver, *Biomed. Mass Spectrom. 12*, 274–287.
88. Lin, S., Shaler, T. A., and Becker, C. H. (2006) Quantification of intermediate-abundance proteins in serum by multiple reaction monitoring mass spectrometry in a single-quadrupole ion trap, *Anal. Chem. 78*, 5762–5767.
89. Anderson, L., and Hunter, C. L. (2006) Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins, *Mol. Cell. Proteomics 5*, 573–588.
90. Zahedi, R. P., Sickmann, A., Boehm, A. M., Winkler, C., Zufall, N., Schönfisch, B., Guiard, B., Pfanner, N., and Meisinger, C. (2006) Proteomic analysis of the yeast mitochondrial outer membrane reveals accumulation of a subclass of preproteins, *Mol. Biol. Cell 17*, 1436–1450.
91. Dunkley, T. P. J., Hester, S., Shadforth, I. P., Runions, J., Weimar, T., Hanton, S. L., Griffin, J. L., Bessant, C., Brandizzi, F., Hawes, C., Watson, R. B., Dupree, P., and Lilley, K. S. (2006) Mapping the Arabidopsis organelle proteome, *Proc. Natl. Acad. Sci. U.S.A. 103*, 6518–6523.
92. Foster, L. J., de Hoog, C. L., Zhang, Y., Zhang, Y., Xie, X., Mootha, V. K., and Mann, M. (2006) A mammalian organelle map by protein correlation profiling, *Cell 125*, 187–199.
93. Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K., Lam, Y. W., Steen, H., Mann, M., and Lamond, A. I. (2002) Directed proteomic analysis of the human nucleolus, *Curr. Biol. 12*, 1–11.
94. Andersen, J. S., Lam, Y. W., Leung, A. K., Ong, S. E., Lyon, C. E., Lamond, A. I., and Mann, M. (2005) Nucleolar proteome dynamics, *Nature 433*, 77–83.
95. Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., and O'Shea, E. K. (2003) Global analysis of protein localization in budding yeast, *Nature 425*, 686–691.
96. Beausoliel, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villén, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) large-scale characterization of HeLa cell nuclear phosphoproteins, *Proc. Natl. Acad. Sci. U.S.A. 101*, 12130–12135.
97. Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X.-M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells, *Nat. Biotechnol. 23*, 94–101.
98. Nousiainen, M., Silljé, H. W., Sauer, G., Nigg, E. A., and Körner, R. (2006) Phosphoproteome analysis of the human mitotic spindle, *Proc. Natl. Acad. Sci. U.S.A. 103*, 5391–5396.
99. Moser, K., and White, F. M. (2006) Phosphoproteomic analysis of rat liver by high capacity IMAC and LC-MS/MS, *J. Proteome Res. 5*, 98–104.
100. Gruhler, A., Olsen, J. V., Shabaz, M., Mortensen, P., Færgemann, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway, *Mol. Cell. Proteomics 4*, 310–327.
101. Pandey, A., Podtelejnikov, A. V., Blagoev, B., Bustelo, X. R., and Mann, M. (2000) Analysis of receptor signaling pathways by mass spectrometry: identification of Vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors, *Proc. Natl. Acad. Sci. U.S.A. 97*, 179–184.
102. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Poepstorff, P., and Jørgensen, J. D. (2005) highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns, *Mol. Cell. Proteomics 4*, 873–886.
103. Tholey, A., Reed, J., and Lehmann, W. D. (1999) Electrospray tandem mass spectrometric studies of phosphopeptides and phosphopeptide analogues, *J. Mass Spectrom. 34*, 117–123.
104. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, *Proc. Natl. Acad. Sci. U.S.A. 101*, 9528–9533.
105. McLachlin, D. T., and Chait, B. T. (2003) Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides, *Anal. Chem. 75*, 6826–6836.
106. Tseng, H.C., Ovaa, H., Wei, N. J., Ploegh, H., and Tsai, L. H. (2005) Phosphoproteomic analysis with a solid-phase capture-release-tag approach, *Chem. Biol. 12*, 769–777.

107. Tao, W. A., Wollscheid, B., O'Brien, R., Eng, J. K., Li, X. J., Bodenmiller, B., Watts, J. D., Hood, L., and Aebersold, R. (2005) Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry, *Nat. Methods 2*, 591–598.
108. Blagoev, B., Ong, S.-E., Kratchmarova, I., and Mann, M. (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics, *Nat. Biotechnol. 22*, 1139–1145.
109. Zheng, H., Hu, P., Quinn, D. F., and Wang, Y. K. (2005) Phosphotyrosine proteomic study of interferon α signaling pathway using a combination of immunoprecipitation and immobilized metal affinity chromatography, *Mol. Cell. Proteomics 4*, 721–730.
110. Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J., Rush, J., Lauffenburger, D. A., and White, F. M. (2005) time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules, *Mol. Cell. Proteomics 4*, 1240–1250.
111. Cantin, G. T., Venable, J. D., Cociorva, D., and Yates, J. R. (2006) Quantitative phosphoproteomic analysis of the tumor necrosis factor pathway, *J. Proteome Res. 5*, 127–134.
112. Carr, S. A., Annan, R. S., and Huddleston, M. J. (2005) Mapping posttranslational modifications of proteins by MS-based selective detection: application to phosphoproteomics, *Methods Enzymol. 405*, 82–115.
113. Williamson, B. L., Marchese, J., and Morrice, N. A. (2006) Automated identification and quantification of protein phosphorylation sites by LC/MS on a hybrid triple quadrupole linear ion trap mass spectrometer, *Mol. Cell. Proteomics 5*, 337–346.