Acc. Chem. Res. **2003,** *36,* 453−461

The Characterization of Protein Post-Translational Modifications by Mass Spectrometry

REBECCA E. SCHWEPPE,[†] CLAIRE E. HAYDON,[†] TIMOTHY S. LEWIS,[†,‡] KATHERYN A. RESING,[†] AND NATALIE G. AHN*[,†,§]

Department of Chemistry and Biochemistry and Howard Hughes Medical Institute, University of Colorado, Boulder, Colorado 80309

Received November 14, 2002

ABSTRACT

Most biological processes are regulated by post-translational modifications of proteins, and conditions that disrupt the regulation of such events can lead to disease. In the past decade, the identification and characterization of covalent modifications have been driven by advances in mass spectrometry. Here, we discuss current mass spectrometric and proteomic approaches for the identification of proteins and their covalent modifications, and we highlight high-throughput strategies for comprehensive analysis of cell proteomes.

Introduction

Post-translational modifications of proteins are critical to protein function and activity. Over 200 distinct covalent modifications have been reported, with phosphorylation, ADP-ribosylation, nitrosylation, glycosylation, lipid modification, ubiquitination, and acetylation being among the most common (for a comprehensive listing, see http://www.abrf.org/index.cfm/dm.home?AvgMass=all). Phosphorylation is one of the most important and abundant, with more than 30% of proteins being modified by the covalent attachment of phosphate.[1] Most likely, all proteins will turn out to be post-translationally modified ("Walsh's rule"), and it is clear that covalent modifications of proteins are essential for numerous cellular functions.

A variety of covalent modifications are involved in the propagation of signaling pathways from the plasma membrane to the nucleus. For example, tyrosine phosphorylation of receptor tyrosine kinases at the membrane is required for receptor dimerization, activation, and subsequent recruitment of adaptor molecules and other protein components within signaling complexes. Serine and threonine phosphorylation promotes association and activation of downstream protein kinases and, along with ubiquitination and sumoylation, regulate the localization of protein complexes to and from the nucleus. The covalent modification of transcription factors in the nucleus by serine/threonine kinases, acetyltransferases, methyltransferases, or ubiquitin ligases results in the activation or repression of gene expression. Further regulation of gene expression is achieved through the modification of histones by acetylation, methylation, phosphorylation, and ubiquitination. Thus, a typical signaling pathway requires many distinct post-translational modifications to elicit specific responses. Given the central role of post-translational modifications in cellular function, the proteome-wide identification of protein covalent modifications is an important goal, and mass spectrometry has emerged as a leading technology to achieve this objective. This review outlines approaches for and provides examples of analyzing protein post-translational modifications by mass spectrometry, with emphasis on phosphoproteomic strategies.

Traditionally, the covalent modification of proteins by phosphorylation has been studied using metabolic labeling with [^{32}P]-orthophosphate.[2,3] Radiolabeled phosphoproteins are purified and characterized after enzymatic digestion by phosphopeptide mapping and phosphoamino acid analysis. First, the sequence of the phosphopeptide is determined by Edman degradation after separation and isolation by reversed-phase high performance liquid chromatography (HPLC). Here, the peptide is chemically sequenced by sequential removal and detection of each amino-terminal residue following derivatization of the amino-terminal amino acid with phenylisothiocyanate and subsequent treatment with acid to produce a phenylthiohydantoin derivative. The specific site of phosphorylation within the labeled phosphopeptide is then identified by correlating the release of ^{32}P during solid-phase sequencing with the amino acid sequence.[3] This technique is advantageous when working with small amounts of sample; in vivo radiolabeling is less successful with peptides containing multiple phosphorylated residues that are partially occupied, or large peptides where the occupied site is >10 residues from the N-terminus, given that the phosphate is slightly labile during the chemical steps and a small amount may be removed with each cycle.

Peptide Sequencing by Mass Spectrometry

Mass spectrometry has significantly advanced analytical methods for protein identification and analysis. Soft

Rebecca Schweppe received her Ph.D. in Biochemistry from the University of Colorado HSC (2000). She is an HHMI fellow of the Life Sciences Research Foundation at the University of Colorado.

Claire Haydon obtained her Ph.D. from the University of Dundee (2002). She is an HHMI postdoctoral associate at the University of Colorado.

Timothy Lewis obtained his Ph.D. in Chemistry/Biochemistry at the University of Colorado. He is a scientist at CEPTYR, Inc.

Katheryn Resing received a Ph.D. in Biochemistry from the University of Washington (1983), where she pursued her postdoctoral studies. She has been a faculty member in the Department of Chemistry/Biochemistry at the University of Colorado since 1992.

Natalie Ahn received a Ph.D. in Chemistry from the University of California, Berkeley (1985), followed by postdoctoral studies at the University of Washington. She has been a faculty member in the Department of Chemistry/Biochemistry at the University of Colorado since 1992.

* Corresponding author. Address: University of Colorado, Department of Chemistry and Biochemistry, Campus Box 215, Boulder, CO 80309. Telephone: 303-492-4799. FAX: 303-492-2439. E-mail: Natalie.Ahn@colorado.edu.

[†] Department of Chemistry and Biochemistry.

[‡] Current address: CEPTYR, Inc., 22215 26th Avenue SE, Bothell, WA 98021.

[§] Howard Hughes Medical Institute.

10.1021/ar020143l CCC: $25.00
Published on Web 05/22/2003

ionization techniques such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) along with improved ion detectors provide accurate mass determination along with enhanced sensitivity and capabilities for high-throughput analysis. In ESI, a fine spray of charged droplets produced through a high voltage needle is desolvated in the high vacuum chamber of the mass spectrometer. Desorption of ions produces multiply protonated forms observable in the positive ion mode, as well as multiply deprotonated forms observable in negative ion mode.[4] Protein or peptide ions are transmitted into a mass analyzer and manipulated by electric fields that allow resolution of mass-to-charge (m/z) ratio. In MALDI, samples are cocrystallized with an organic matrix that is volatile when excited by pulses of laser light.[5] Upon laser excitation, the matrix and sample are desorbed, and primarily singly charged ions are produced. MALDI is often coupled with time-of-flight (TOF) mass analyzers in which the ion m/z is determined by the time required to travel from the ion source to the detector.

To identify proteins by mass spectrometry, proteins are digested in solution after purification or digested in situ from stained gels, using a specific protease, such as trypsin. The resulting peptides are desalted and subjected to MS analysis. Direct measurements of mass (often using MALDI-TOF) allows protein identification by peptide mass fingerprinting, where the observed peptide masses are compared to theoretical peptide masses of protein database entries. The development of computer programs for identifying proteins from peptide mass spectral data represented a key advance in this area.[6−9]

If necessary, peptide sequences are confirmed by fragmentation (MS/MS) using collision-induced dissociation (CID or field resonance-induced dissociation).[10] For example, in either a tandem quadrupole or hybrid qTOF instruments, the first mass analyzer scans a wide mass range and, by manipulating the quadrupole field, selects an ion referred to as a "precursor" ion which then undergoes CID in the second quadrupole. Fragment ion spectra are then scanned in either a third quadrupole or a TOF mass analyzer. MS/MS analysis with ion trap instruments utilize field manipulation to allow only the mass selected ion to remain in the trap, which is resonantly excited to induce fragmentation. Fragment ions of each m/z range are then analyzed by progressively destabilizing each mass range so that they sequentially exit the trap to the detector.

During CID, peptides are preferentially fragmented along the peptide backbone. Six types of fragment ions that may be generated include a-, b-, and c-series ions, containing the peptide N-terminus, and x-, y-, and z-series ions, containing the peptide C-terminus (Figure 1A).[11] Any position in the peptide backbone can be fragmented, including alkyl-carbonyl bonds (CHR−CO), producing a and x ions, peptide-amide bonds (CO−NH), producing b and y ions, and amino-alkyl bonds (NH−CHR), producing c and z ions. Cleavage of amide bonds is most common, generating b- and y-type ions. Additional fragment ions resulting from the loss of water, carbon monoxide, or ammonia can also be generated during CID and facilitate the identification of specific amino acids within a spectrum.

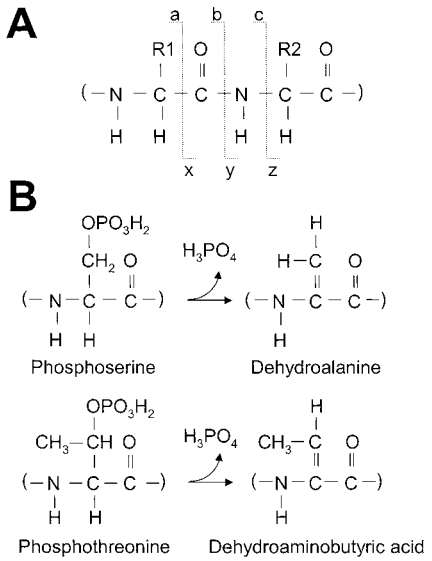

FIGURE 1. Fragmentation chemistries observed during MS/MS. (A) Nomenclature of product ions. CID induces cleavage at different bonds in the peptide backbone. The a-, b-, and c-type ions represent fragment ions containing the peptide N-terminus while the x-, y-, and z-ions represent ions containing the peptide C-terminus. (B) β-Elimination of phosphoserine and phosphothreonine. During CID, phosphoserine, and phosphothreonine undergo β-elimination to dehydroalanine and dehydroaminobutyric acid, respectively. Such reactions also occur in solution under alkaline conditions.

Examples of Covalent Modifications Analyzed by Mass Spectrometry

Phosphorylation. Phosphopeptides are detected by increases in peptide mass characteristic of covalent modification by phosphate (HPO_3), which yields +80 or +160 Da, respectively, for mono- or diphosphorylated species. Peptide mass differences of −80 Da following treatment of proteins with phosphatase also facilitates the identification of phosphopeptides.

An example of MS/MS analysis of an unphosphorylated tryptic peptide and its diphosphorylated form is shown in Figure 2. The MS/MS spectrum of the unphosphorylated peptide (MH_3^{+3} = 808.3 Da), derived from the activation loop of MAP kinase kinase-1 (MAPKK1, residues 206−227) shows a series of y ions (y_1−y_{12}) and b ions (b_2−b_{17}) which report the indicated sequence (Figure 2A). Phosphorylation sites are identified by scanning for daughter ions with masses 80 Da larger than the corresponding unphosphorylated forms. Specifically, the parent ion (MH_3^{+3} = 861.6 Da) generated ions y_6 + 80 through y_9 + 80 Da as well as y_{10} + 160 through y_{11} + 160 Da, indicating phosphorylation at residues Ser-218 and Ser-222 (Figure 2B).

The identification of ions with masses 18 Da less than the unphosphorylated forms are also observed, representing the neutral loss of phosphoric acid (H_3PO_4, 98 Da) from phosphoserine or phosphothreonine (Figure 2).Neutral loss occurs through gas phase β-elimination of H_3PO_4 from phosphoserine or phosphothreonine residues, which are converted to dehydroalanine or dehydroami-

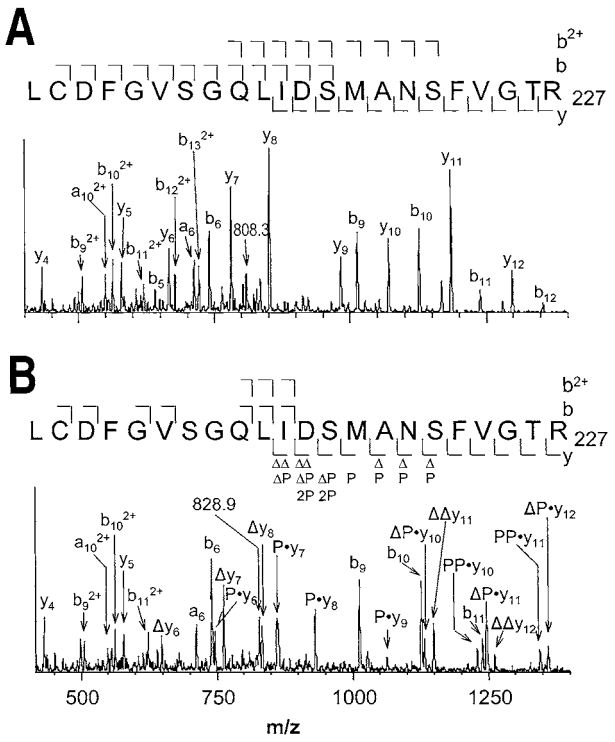

FIGURE 2. MS/MS identification of phosphorylation sites within the activation loop of MAPKK1. (A) MS/MS Spectrum corresponding to unphosphorylated MAPKK1 (tryptic peptide 206−227). Indicated are the parent ion (MH_3^{+3} = 808.3 Da) as well as the observed b ([) and y (]) cleavage products of this peptide. (B) MS/MS spectrum corresponding to the diphosphorylated form of peptide 206−227. Indicated is the parent ion (MH_3^{+3} = 828.9 Da). Fragment ions with mass +80 or +160 Da from unmodified are respectively denoted with "P" or "PP". β-Eliminated ions with mass −18 or −36 Da from unmodified (−98 or −196 Da from phosphorylated) are denoted by "Δ" or "ΔΔ". Fragment ions that are initially diphosphorylated but β-eliminated at one site (+62 from unmodified) are denoted by "ΔP". Figure modified from Resing and Ahn (1995).[49]

nobutyric acid, respectively (Figure 1B). In the example, fragment ions $y_6 - 18$ through $y_8 - 18$ Da and $y_{11} - 36$ and $y_{12} - 36$ Da represent neutral loss of one and two phosphoric acid groups, respectively. Finally, ions $y_{10} + 62$ through $y_{12} + 62$ Da are also detected, representing β-elimination of diphosphorylated fragment ions (+160−98 Da). Thus, phosphorylation at Ser-218 and Ser-222 can be unequivocally identified by careful analysis of complex products of gas-phase chemistry.

Phosphopeptides can also be identified using "neutral loss scanning" in tandem quadrupole mass spectrometers. This method uses CID in positive ion mode to induce β-elimination of H_3PO_4 from phosphoserine- or phosphothreonine-containing peptides. Ions are scanned in parallel in two mass analyzers (e.g., first and third quadrupoles) with an offset corresponding to loss of 98 Da, such that the only ions to reach the detector are those that lose 98 Da upon fragmentation. After scanning parent ions over a wide mass range in the first mass analyzer, products generated by CID which show a loss of 98 Da in a second mass analyzer are selected for sequencing.[12] Phosphotyrosine residues cannot be identified by neutral loss scanning due to stability of the β-protons in the benzene ring which minimize β-elimination.

"Precursor ion scanning" uses tandem MS in negative ion mode to identify phosphopeptides.[13,14] This method scans ions over a wide mass range in the first mass analyzer, followed by CID. When performed under alkaline conditions, negative ions of $m/z = 79$ formed by release of PO_3^- from phosphopeptides can be optimally detected in the first mass analyzer. Each corresponding parent ion can then be selected and analyzed by MS/MS after switching to positive ion mode. Precursor ion scanning can also be used to identify phosphotyrosine-containing peptides by scanning in positive ion mode for the appearance of the phosphotyrosine immonium ion, which corresponds to the internal fragment ion of phosphotyrosine minus its C-terminal carbonyl ($m/z = 216.04$).[15] Because phosphotyrosine-containing peptides are identified in positive ion mode, MS/MS sequencing of the parent ion can be performed simultaneously.

The high selectivity of neutral loss or precursor ion scanning enables the identification of phosphopeptides in complex mixtures. Although phosphopeptides can be sequenced at femtomole levels, the identification of the occupied residues often remains a challenge due to lower abundance of phosphorylated versus unphosphorylated peptides in a digest mixture. While neutral loss or precursor ion scanning generally requires larger quantities of sample, these procedures simplify the analysis by selecting only the phosphorylated parent ions from complex mixtures.

Acetylation and Methylation. The example of reversible histone acetylation and deacetylation represents a well-characterized, evolutionarily conserved post-translational modification that plays an important role in transcriptional regulation, DNA repair and replication. Other modifications, including phosphorylation, methylation, ADP-ribosylation, and ubiquitination, have been shown to be important regulators of histone function. For example, phosphorylation of histone H3 at Ser-10 has been shown to be important for gene activation and chromosome condensation during mitosis.[16] Methylation of histones H3 and H4 at arginine residues generally results in transcriptional activation, while methylation at lysine can lead to repression or activation, depending on the state of methylation.[17−20]

Mass spectrometric approaches have proven advantageous in identifying covalent modifications on histones with respect to high sensitivity and speed. Recent MALDI-TOF and tandem MS studies have characterized the level and order of lysine acetylation of histone H4 isolated from HeLa cells.[21] In a separate study, histone H3 modifications were identified using MALDI-TOF by monitoring characteristic immonium ions to rapidly distinguish between methylated versus acetylated forms of histone H3.[22]

Histones provide a good example of using intact protein mass analysis to directly and rapidly monitor post-translational modifications. Our laboratory has previously characterized covalent modifications on intact histones isolated from human erythroleukemia cell chromatin.[23]

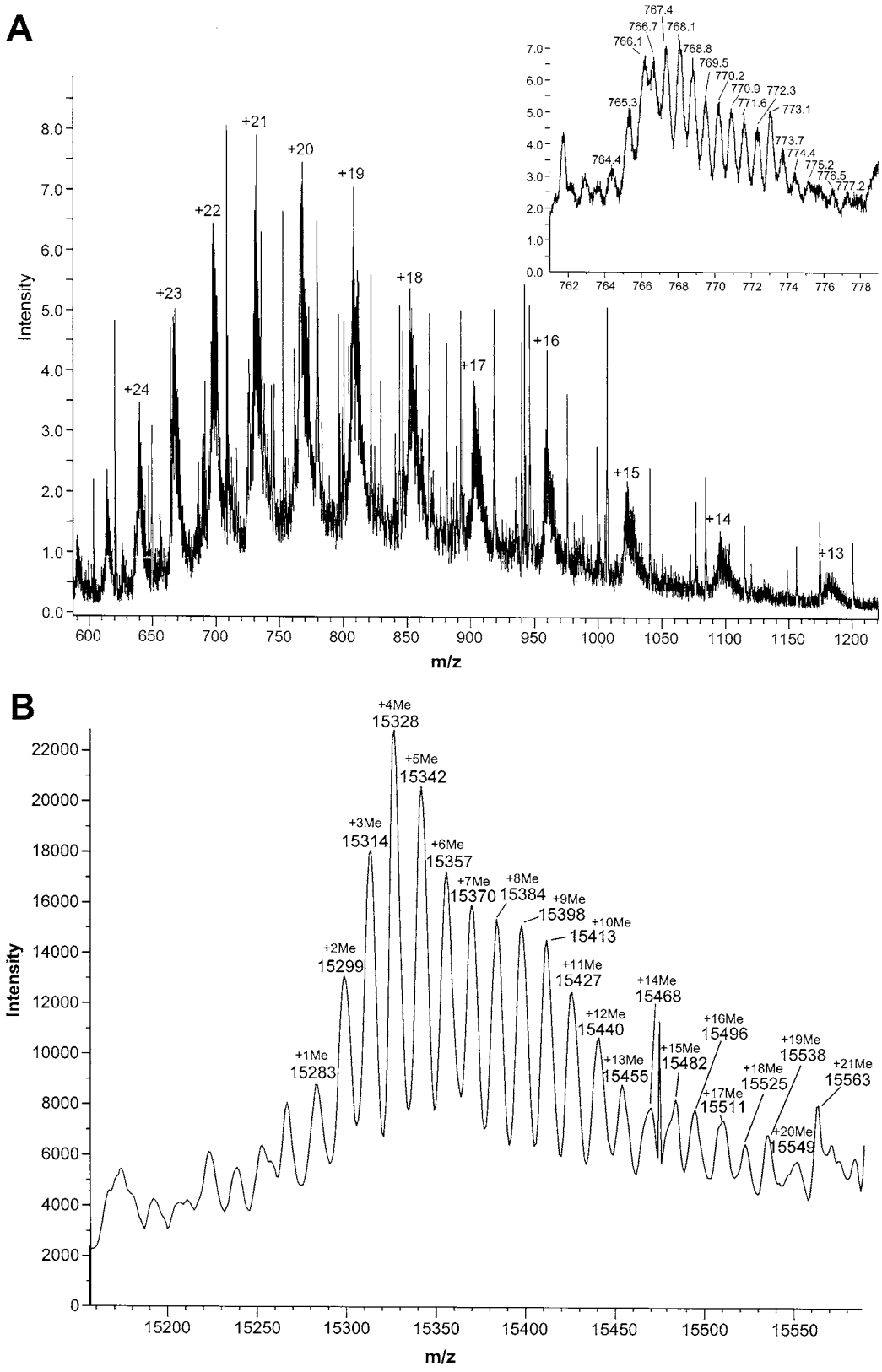

FIGURE 3. MS analysis of modified histone H3. Histones (30 μg) were purified from K562 cells,[23] acidified, and analyzed by ESI/LC/MS. Reversed-phase columns were packed with POROS 20 R1 perfusion chromatography resin, and spectra were collected using an ABI Q-Star Pulsar mass spectrometer. (A) LC/MS spectrum of histone H3. Ions corresponding to different charged species are shown, where each species provides an independent measurement of mass. Inset: an expanded view of the +20 charged ion shows additional peaks reflecting covalent modifications. (B) Deconvolution averaging peaks corresponding to differentially modified forms of histone H3. The deconvolution plots were derived from the ion spectra for histone H3 using Bayesian Protein Reconstruction (ABI Analyst QS software). Each peak is labeled by intact protein mass and predicted stoichiometry of methylation.

Figure 3A shows an MS spectrum of histone H3, where observed peaks represent different charge forms of the full length protein. Close examination of individual charge forms (e.g., +20 ion; Figure 3A inset) reveals side peaks which reflect covalently modified forms. A deconvolution plot averaging over all charge states shows the masses of modified forms of histone H3 (Figure 3B). Masses increasing by approximately +14 Da increments are consistent with arginine- or lysine-methylated species. Methylated forms with stoichiometry up to 21 are indicated, consistent with known Lys-4, -9,- 27, -36, and -79, which can be modified with up to three methyl groups, and Arg-2, -17, and -26, which can be modified with up to two groups.[24,25]

Myristoylation. *N*-Myristoylation is a co- or post-translational modification in which myristic acid (C14:0) is attached through an amide linkage to the amino-terminal glycine of eukaryotic and viral proteins containing the consensus sequence Met-Gly-X-X-X-Ser/Thr-.[26] *N*-Myristoylation occurs on a variety of proteins, resulting in diverse functional consequences through the regulation of protein-membrane and protein−protein interactions.

MS-based techniques have succeeded in identifying sites of myristoylation in a variety of proteins, including a previously uncharacterized myristoylation site on the human rhinovirus protein, VP4.[27] Mass spectrometry indicated the observed mass of VP4 was 212 Da larger than the expected mass, consistent with modification by myristate, and confirmed the N-terminal site of myristoylation by MS/MS.

A second example includes members of the ADP-ribosylation factor (ARF) family of small GTP-binding proteins, which consists of six highly conserved proteins, making it difficult to distinguish ARFs by immunological methods. Additionally, identification by Edman sequencing has been difficult because the unique sequences that distinguish various ARFs are clustered at the N-terminus, which is blocked by *N*-myristoylation.

Figure 4 shows the identification and analysis of a purified bovine ARF, named GGBF*.[28] GGBF and GGBF* (GTP-dependent Golgi binding factor) are two proteins of similar size (21 kDa) that were co-purified from brain and both found to be immunologically related to the ARF family, although each exhibited distinct specific activities.[29] Figure 4 shows the MS spectra and deconvolution plot of GGBF*, indicating a single mass of 20 678 Da, consistent with the size of myristoylated ARF3 (Figure 4B). In contrast, GGBF showed a single mass of 20 774 Da, consistent with myristoylated ARF1 (not shown).[28]

To confirm the respective assignments of GGBF and GGBF* to ARF1 and ARF3, LC/MS/MS of trypsin digests was performed.[28] For GGBF*, the N-terminal peptide (m/z = 1185.9 Da) was a unique identifier. Its MS/MS spectrum is shown in Figure 4C where overlapping sets of b (b_1−b_4) and y (y_5−y_8) ions confirm its identity as ARF3. Additionally, a fragment ion of m/z 211.9 was identified, confirming *N*-myristoylation of ARF3 (Figure 4C). Thus, accurate mass determination of the intact protein in combination with MS/MS sequencing identified and distinguished two highly related proteins and simultaneously mapped sites of lipid modification.

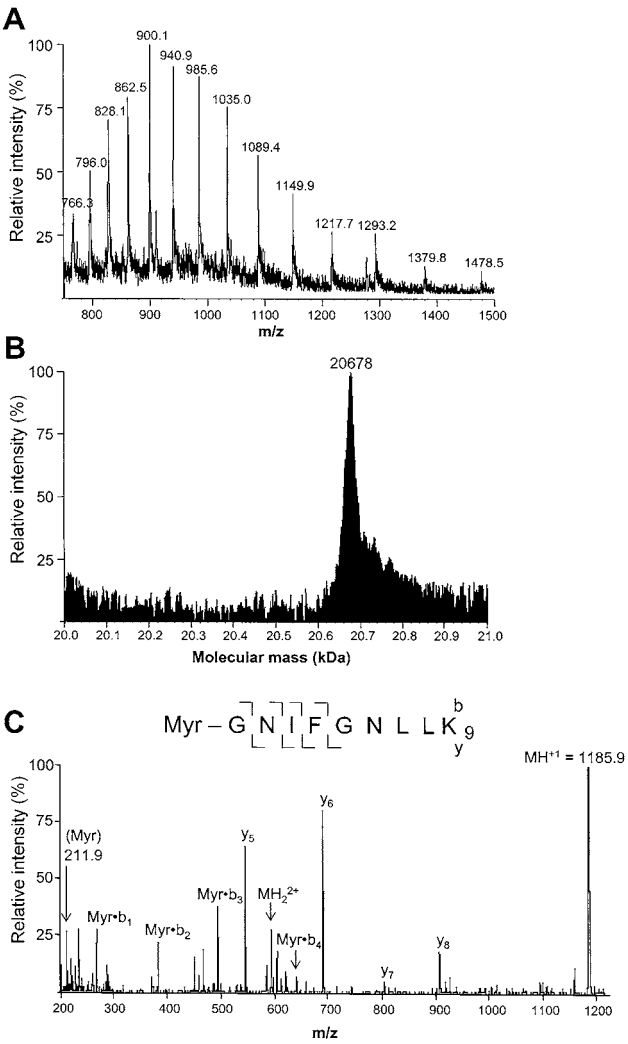

FIGURE 4. MS Identification of ARF3. (A) LC/MS spectrum of ARF3 (GGBF*). Multiply charged ion species are shown. (B) Deconvolution plot showing a single symmetrical peak centered at 20 678 ± 3 Da. The observed mass of the full-length protein is consistent with ARF3 that is demethionylated and myristoylated at the N-terminus (predicted mass 20 681 Da). (C) MS/MS spectrum of the parent ion MH+1 = 1185.9 Da, corresponding to the N-terminal peptide of ARF3. The b_1−b_4 fragment ions are consistent with the N-terminal peptide of ARF3 containing an N-terminal modifying myristate (Myr), which is also observed (predicted m/z = 212.2 Da). Modified from Berger et al. (1995).[28]

Global Analysis of Post-Translational Modifications

The increased sensitivity of mass spectrometers, together with improved methods for in-gel digestion and innovative computer search algorithms for identifying proteins from mass spectral data, allows identification of femtomole levels of proteins recovered after SDS−PAGE separation. This has enabled functional proteomic strategies that combine 2D gel analysis with mass spectrometry for monitoring global protein changes in response to signal transduction pathways.[30,31] Proteins in complex protein

mixtures are resolved on 2D gels and stained gels are examined for changes in protein migration with respect to pI or mass. Proteins that change in intensity are excised, digested in-gel with protease, and identified by peptide mass fingerprinting and MS/MS sequencing.

Using this approach, we have identified 25 protein targets downstream of the MAP kinase cascade in human K562 cells, including 20 new targets of this signaling pathway.[31] Importantly, the analysis showed that 14 of the proteins were obviously regulated by post-translational modifications based on their altered pI mobilities on 2D gels. Most of the remaining proteins were also likely to be regulated covalently because quantitation of intensity versus time showed nearly all of the targets responding within the first 20 min of cell stimulation. The results indicate that 2D gel strategies can be used to screen post-translationally regulated proteins responsive to activation of signaling pathways.

An example shows the altered 2D gel mobility of the tubulin binding protein, stathmin, upon cell stimulation by PMA (Figure 5A), which is suppressed by inhibitors of MAPKK1/2 (not shown). Cell stimulation resulted in the reduced abundance of a form present under basal conditions (U-88), and the subsequent increased abundance of two forms with more acidic pI (U-87 and U-86), all three of which were identified by peptide mass fingerprinting as stathmin (not shown). U-88 yielded tryptic peptides with mass 1326.57 and 1388.67 Da, which were respectively converted in U-86 to forms at 1406.44 and 1468.48 Da (Figure 5B). In each case the +80 Da increase in mass revealed modification by phosphorylation. The MS/MS of the 1468.5 Da phosphopeptide ion showed the series of y ions (y_2-y_9) and b ions (b_5-b_{10}) (Figure 5C). Increased masses of +80 Da for y_3-y_9 (indicating phosphorylation) and decreased masses of −18 Da for y_7, b_9, and b_{10} (indicating dehydration) allowed assignment of the phosphorylation site to Ser-25. These studies demonstrate the success of 2D gels in screening for functional changes in covalent modification in response to signaling pathways.

To date, 2D gel approaches can detect up to ~6000 different protein features from cell lysates, by separating proteins using several narrow-range pI intervals which overlap to cover the full pH range.[32] However, the utility of 2D gels is limited with respect to staining sensitivity and resolution as well as poor recovery of insoluble proteins. Thus, representation is often precluded for proteins which are low in abundance (<10 000 copies/cell), high in mass (>100 kDa), or membrane-associated. Thus, although 2D gels provide a rich source of information about protein abundance, mass, and pI, new technologies are required to comprehensively characterize the proteome.

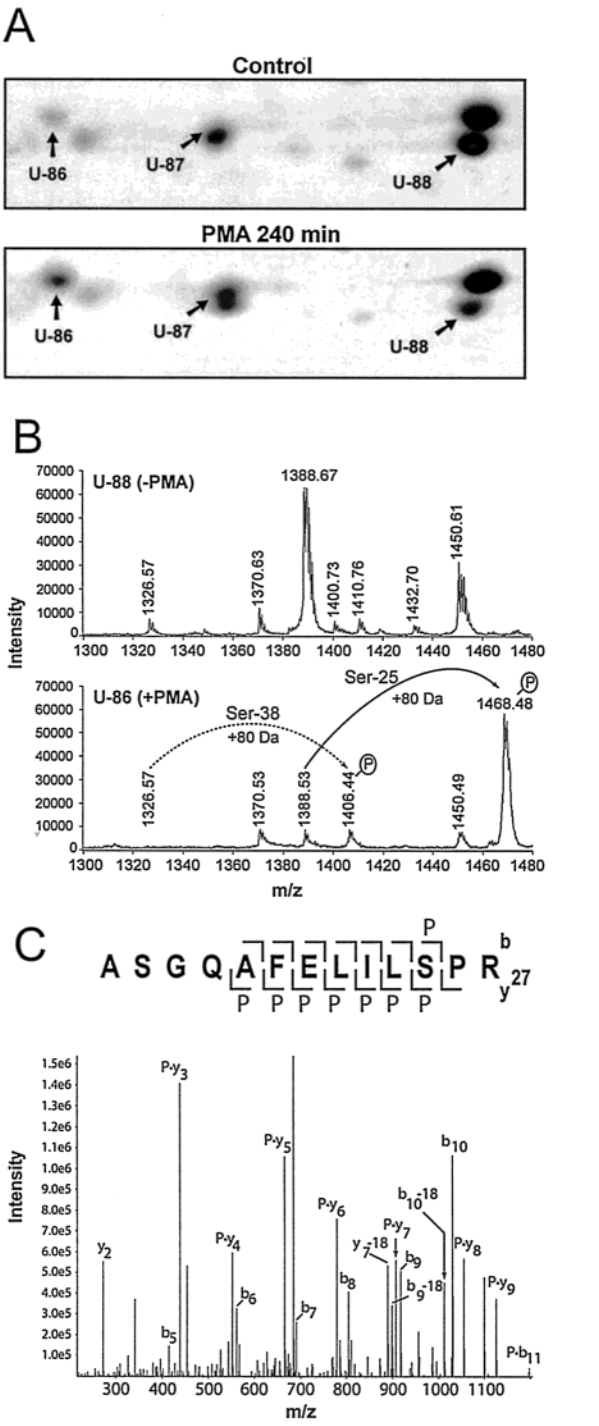

FIGURE 5. 2D gel and MS analysis of stathmin. (A) 2D gel analysis of K562 extracts treated with or without PMA. 2D gels of K562 extracts treated ±PMA and visualized by silver staining show functional changes in unknown proteins U-88, U-87 and U-86. **B)** MALDI-TOF spectrum of U-86 vs U-88. The shift in U-86 peptide masses 1326.57 to 1406.44 and 1383.53 to 1468.48 by +80 Da is shown (P). **C)** MS/MS spectrum of the MH_2^{+2} = 734.8 Da parent ion, corresponding to the peptide with mass 1468.5 Da in panel B. The series of b (⌐) and y (⌙) fragment ions is shown. The peptide sequence corresponds to residues 15−27 of human stathmin. The observed masses of ions y_3-y_9 and b_{11} were increased by 80 Da, indicating phosphorylation of two peptides following PMA stimulation, which reflects modification at Ser-25 and Ser-27, respectively (P). Masses of ions y_7, b_9, and b_{10} were 18 Da less than expected, reflecting dehydration of these fragments.

New Directions for Profiling Covalent Modifications

Protein Identification by Multidimensional LC/MS/MS. Automated LC/MS/MS techniques are being exploited in alternative approaches to high throughput functional

proteomics. Typically, peptides are separated by reversed-phase HPLC followed by in-line analysis with mass spectrometry. Multiple ions can be selected for fragmentation every several seconds, allowing several hundred MS/MS spectra to be collected per LC/MS run. However, because ion selection is based on intensity, preferential selection of peptides derived from higher abundance proteins may limit the dynamic range. Thus, strategies are required to simplify the peptide complexity in each MS/MS cycle.

A promising approach is multidimensional liquid chromatography interfaced with MS/MS. Link et al. (1999) reported direct analysis of digests of large protein complexes (DALPC), by fractionating peptides on the basis of charge and hydrophobicity prior to direct MS analysis.[33] Using this approach, digested yeast ribosomal proteins were separated on tandem columns packed with strong cation exchange (SCX) reversed-phase (RP) resins. After loading the peptides onto the SCX column, peptides were step-eluted with a salt gradient, adsorbed to the RP column, eluted with acetonitrile, and analyzed in-line with MS/MS. A variation of this method, denoted multidimensional protein identification technology (MudPIT), uses SCX and RP resins tandemly packed into a single column.[34,35]

MS/MS spectra are typically analyzed using automated database search engines, which assign sequences and report scores based on similarities between observed and theoretical fragmentation. Using DALPC or MudPIT, novel components within yeast ribosomes and mitotic exit complexes were identified, as well as 1484 proteins from yeast protein lysates.[33,34] Recent studies have characterized the rice and *Plasmodium falciparum* proteomes in which 2528 and 2415 unique proteins were identified, respectively.[36,37] Another approach, termed accurate mass tag analysis, takes advantage of high mass accuracies (1 ppm) achievable by Fourier transform ion cyclotron resonance.[38−40] Accurate mass tag analysis of the proteome of *Deinococcus radiodurans*, has revealed 1910 proteins, representing 61% of predicted open reading frames.[41] In summary, multidimensional chromatography and identification of accurate mass tags reduce sample complexity to enable direct analysis of a wide range of proteins represented in a given proteome.

Identification of post-translationally modified peptides in complex mixtures remains a challenge with high throughput approaches to functional proteomics. Initial success in this area has been reported by combining multiproteolytic cleavage with MudPIT.[42] Using this approach, 18 modified residues were identified in a simple mixture of five Cdc2p-interacting proteins, and 73 modified residues were identified within 11 forms of Crystallin from lens tissue.[42] The strategy works best when the number of protein sequences searched is relatively small. Further optimization of this technique to search larger databases will provide a powerful tool for comprehensively mapping protein post-translational modifications.

Enriching Covalently Modified Peptides and Proteins. *Phosphoproteomics.* Several selection techniques have been developed to facilitate the identification of phosphoproteins present in complex mixtures. A classic technique is immobilized metal affinity chromatography (IMAC), which utilizes Fe^{3+}- or Ga^{3+}-chelated solid supports to selectively bind and enrich phosphopeptides. However, these resins also bind peptides through metal interactions with acidic residues and/or electron donors. Ficarro et al reported an improvement of this method, in which peptides are methylesterified prior to IMAC, reducing nonspecific binding by blocking acidic residues.[43] In combination with LC/ESI/MS, over 200 phosphopeptides were detected from tryptic digests of *Saccharomyces cerevisiae* extracts.[43]

Alternative approaches use new chemical protocols for selectively modifying phosphopeptides or phosphoproteins within complex mixtures. Chemically modified peptides are then enriched by covalent or high affinity avidin−biotin coupling to immobilized supports, allowing stringent washing to remove nonphosphorylated forms. One recent enrichment technique involves β-elimination of phosphate, followed by addition of ethanedithiol. The free sulfhydryl group formed can be coupled to biotin, allowing subsequent purification of the adduct on NeutrAvidin resin.[44] A second protocol involves N- and C-terminal protection, carbodiimide condensation with cystamine, and solid-phase capture of resulting free sulfhydryls on glass beads coupled to iodoacetic acid.[45] Both methods successfully isolated serine- and threonine-containing phosphopeptides from whole cell yeast lysates, although phosphotyrosyl peptides were absent and only phosphopeptides from high abundance proteins were represented.[45]

Glycomics. Enrichment methods provide a good strategy for comprehensively surveying other covalent modifications, provided that affinity supports are available. Such methods are being applied to the analysis of glycosylation, which occurs in eukaryotic cells in the form of *N*- and *O*-linked glycosylation as well as GPI anchoring. Glycosylation is involved in many cellular functions, including differentiation, development and morphogenesis. Glycans are composed of different combinations of saccharides, which form complex, branched structures. Due to the numerous types of glycans synthesized within a cell, the global analysis of glycoproteins in a high throughput manner has been challenging.

Gel-based techniques have been successful in characterizing the glycosylation of individual proteins in addition to the types of oligosaccharides present.[46,47] Following in-gel digestion, oligosaccharide sequence information can be obtained by MALDI-TOF MS or ESI/MS. However, potentially more comprehensive are direct MS approaches, such as "glyco-catch," in which sequential lectin affinity columns are used to enrich specific glycoproteins followed by protease digestion. Glycopeptides are recaptured and eluted using the same lectin columns, resolved by trimethylsilyl reversed phase HPLC, and analyzed by mass using MALDI-TOF or Edman degradation. Information for databasing includes protein sequence, glycopeptide mass, 2D-gel mobility, and relative affinity for various

lectins measured by frontal affinity chromatography. Such strategies have been applied to the *Caenorhabditus elegans* "glycome project", whose eventual goal is to identify every glycan produced in this organism.[48] Identification of more than four dozen glycopeptides from *C. elegans* has been achieved, demonstrating proof of principle for more extensive databasing of the worm glycome.

Conclusions

Advancements in mass spectrometry have proven invaluable toward the analysis of post-translational modifications. The continued development of mass spectrometric methods applied to single proteins, along with new methods for high throughput analysis of complex protein mixtures, will provide a framework for the comprehensive analysis of covalent modifications within cells and new insights into basic regulatory mechanisms that control cell function, specificity, and the development of disease.

References

(1) Hubbard, M.; Cohen, P. On target with a new mechanism for the regulation of protein phosphorylation. *Trends Biochem. Sci.* **1993**, *18*, 172−177.

(2) Boyle, W.; van der Geer, P.; Hunter, T. Phosphopeptide mapping and phosphoamino acid analysis by two-dimensional separation on thin-layer cellulose plates. *Methods Enzymol.* **1991**, *201*, 110−149.

(3) Wettenhall, R.; Aebersold, R.; Hood, L. Solid-phase sequencing of 32P-labeled phosphopeptides at picomole and subpicomole levels. *Methods Enzymol.* **1991**, *201*, 186−199.

(4) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*.

(5) Hillenkamp, F.; Karas, M.; Beavis, R.; Chait, B. Matrix-assisted laser desorption/ionization of biopolymers. *Anal. Chem.* **1991**, *63*, 1193A-1203A.

(6) Mann, M.; Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **1994**, *66*, 4390−4399.

(7) Henzel, W.; Billeci, T.; Stults, J.; Wong, S.; Grimley, C.; Watanabe, C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011−5015.

(8) Clauser, K.; Hall, S.; Smith, D.; Webb, J.; Andrews, L.; Tran, H.; Epstein, L.; Burlingame, A. Rapid mass spectrometric peptide sequencing and mass matching for characterization of human melanoma proteins isolated by two-dimensional PAGE. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5072−5076.

(9) Zhang, W.; Chait, B. Profound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal. Chem.* **2000**, *72*, 2482−2489.

(10) Andersen, J.; Svensson, B.; Roepstorff, P. Electrospray ionization and matrix assisted laser desorption/ionization mass spectrometry: powerful analytical tools in recombinant protein chemistry. *Nat. Biotechnol.* **1996**, *14*, 449−457.

(11) Biemann, K. Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods Enzymol.* **1990**, *193*, 445−479.

(12) Covey, T.; Shushan, B.; Bonner, R.; Schroder, W.; Hucho, F. In *Methods in Protein Sequence Analysis*; Gustavsson, A., Ed.; Birkhauser Verlag: Basel, 1991; pp 249−256.

(13) Carr, S.; Huddleston, M.; Annan, R. Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry. *Anal. Biochem.* **1996**, *239*, 180−192.

(14) Wilm, M.; Neubauer, G.; Mann, M. Parent ion scans of unseparated peptide mixtures. *Anal. Chem.* **1996**, *68*, 527−533.

(15) Steen, H.; Kuster, B.; Fernandez, M.; Pandey, A.; Mann, M. Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in the positive ion mode. *Anal. Chem.* **2001**, *73*, 1440−1448.

(16) Cheung, P.; Allis, C.; Sassone-Corsi, P. Signaling to chromatin through histone modifications. *Cell* **2000**, *103*, 263−271.

(17) Chen, D.; Ma, H.; Hong, H.; Koh, S.; Huang, S.; Schurter, B.; Aswad, D.; Stallcup, M. Regulation of transcription by a protein methyltransferase. *Science* **1999**, *284*, 2174−2177.

(18) Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B.; Sun, Z.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C.; Allis, C.; Jenuwein, T. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **2000**, *406*, 593−599.

(19) Nielsen, S.; Schneider, R.; Bauer, U.; Bannister, A.; Morrison, A.; O'Carroll, D.; Firestein, R.; Cleary, M.; Jenuwein, T.; Herrera, R.; Kouzarides, T. Rb targets histone H3 methylation and HP1 to promoters. *Nature* **2001**, *412*, 561−565.

(20) Santos-Rosa, H.; Schneider, R.; Bannister, A.; Sherriff, J.; Bernstein, B.; Emre, N.; Schreiber, S.; Mellor, J.; Kouzarides, T. Active genes are tri-methylated at K4 of histone H3. *Nature* **2002**, *419*, 407−411.

(21) Zhang, K.; Williams, K.; Huang, L.; Yau, P.; Siino, J.; Bradbury, E.; Jones, P.; Minch, M.; Burlingame, A. Histone acetylation and deacetylation. *Mol. Cell. Proteomics* **2002**, *1*, 500−508.

(22) Zhang, K.; Tang, H.; Huang, L.; Blankenship, J.; Jones, P.; Xiang, F.; Yau, P.; Burlingame, A. Identification of acetylation and methylation sites of histone H3 from chicken erythrocytes by high-accuracy matrix-assisted laser desorption ionization-time-of-flight, matrix-assisted laser desorption ionization-postsource decay, and nanoelectrospray ionization tandem mass spectrometry. *Anal. Biochem.* **2002**, *306*, 259−269.

(23) Galasinski, S.; Louie, D.; Gloor, K.; Resing, K.; Ahn, N. Global regulation of post-translational modifications on core histones. *J. Biol. Chem.* **2002**, *277*, 2579−2588.

(24) Zhang, Y.; Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **2001**, *15*, 2343−2360.

(25) Ng, H.; Feng, Q.; Wang, H.; Erdjument-Bromage, H.; Tempst, P.; Zhang, Y.; Struhl, K. Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev.* **2002**, *16*, 1518−1527.

(26) Resh, M. Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochim. Biophys. Acta* **1999**, *1451*, 1−16.

(27) Lewis, J.; Bothner, B.; Smith, T.; Siuzdak, G. Antiviral agent blocks breathing of the common cold virus. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6774−6778.

(28) Berger, S.; Resing, K.; Taylor, T.; Malancon, P. Mass-spectrometric analysis of ADP-ribosylation factors from bovine brain: identification and evidence for homogeneous acylation with the C14:0 fatty acid (myristate). *Biochem. J.* **1995**, *311*, 125−132.

(29) Taylor, T.; Kahn, R.; Melancon, P. Two distinct members of the ADP-ribosylation factor family of GTP-binding proteins regulate cell-free intra-Golgi transport. *Cell* **1992**, *70*, 69−79.

(30) Kanamoto, T.; Hellman, U.; Heldin, C.-H.; Souchelnytskyi, S. Functional proteomics of transforming growth factor-β-stimulated Mv1Lu epithelial cells: Rad51 as a target of TGFβ1-dependent regulation of DNA repair. *EMBO J.* **2002**, *21*, 1219−1230.

(31) Lewis, T.; Hunt, J.; Aveline, L.; Jonscher, K.; Louie, D.; Yeh, J.; Resing, K.; Ahn, N. Identification of novel MAP kinase signaling targets by functional proteomics and mass spectrometry. *Mol. Cell* **2000**, *6*, 1343−1354.

(32) Hoving, S.; Voshol, H.; van Oostrum, J. Towards high performance two-dimensional gel electrophoresis using ultrazoom gels. *Electrophoresis* **2000**, *21*, 2617−2621.

(33) Link, A.; Eng, J.; Schieltz, D.; Carmack, E.; Mize, G.; Morris, D.; Garvik, B.; Yates, J. R., III. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotech.* **1999**, *17*, 676−682.

(34) Washburn, M.; Wolters, D.; Yates, J. R., III Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotech.* **2001**, *19*, 242−247.

(35) Wolters, D.; Washburn, M.; Yates, J. R., III An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73*, 5683−5690.

(36) Koller, A.; Washburn, M.; Lange, B.; Andon, N.; Deciu, C.; Haynes, P.; Hays, L.; Schieltz, D.; Ulaszek, R.; Wei, J.; Wolters, D.; Yates, J. R., III. Proteomic survey of metabolic pathways in rice. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11969−11974.

(37) Florens, L.; Washburn, M.; Raine, J.; Anthony, R.; Grainger, M.; Haynes, J.; Moch, J.; Muster, N.; Sacci, J.; Tabb, D.; Witney, A.; Wolters, D.; Wu, Y.; Gardner, M.; Holder, A.; Sinden, R.; Yates, J. R., III; Carucci, D. A proteomic view of *Plasmodium falciparum* life cycle. *Nature* **2002**, *419*, 520−526.

(38) Smith, R.; Anderson, G.; Lipton, M.; Pasa-Tolic, L.; Shen, Y.; Conrads, T.; Veenstra, T.; Udseth, H. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* **2002**, *2*, 513−523.

(39) Shen, Y.; Zhao, R.; Belov, M.; Conrads, T.; Anderson, G.; Tang, K.; Pasa-Tolic, L.; Veenstra, T.; Lipton, M.; Udseth, H.; Smith, R. Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Anal. Chem.* **2001**, *73*, 1766−1775.
(40) Conrads, T.; Anderson, G.; Veenstra, T.; Pasa-Tolic, L.; Smith, R. Utility of accurate mass tags for proteome-wide identification. *Anal. Chem.* **2000**, *72*, 3349−3354.
(41) Lipton, M.; Pasa-Tolic, L.; Anderson, G.; Anderson, D.; Auberry, D.; Battista, J.; Daly, M.; Frederickson, J.; Hixson, K.; Kostandarithes, H.; Masselon, C.; Markillie, L.; Moore, R.; Romine, M.; Shen, Y.; Stritmatter, E.; Toli, N.; Udseth, H.; Amudhan, V.; Wong, K.; Zhao, R.; Smith, R. Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11049−11054.
(42) MacCoss, M.; McDonald, W.; Saraf, A.; Sadygov, R.; Clark, J.; Tasto, J.; Gould, K.; Wolters, D.; Washburn, M.; Weiss, A.; Clark, J.; Yates, J. R., III Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7900−7905.
(43) Ficarro, S.; McCleland, M.; Stukenberg, P.; Burke, D.; Ross, M.; Shabanowitz, J.; Hunt, D.; White, F. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotech.* **2002**, *20*, 301−305.
(44) Oda, Y.; Nagasu, T.; Chait, B. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotech.* **2001**, *19*, 379−382.
(45) Zhou, H.; Watts, J.; Aebersold, R. A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **2001**, *19*, 375−378.
(46) Rudd, P.; Colominas, C.; Royle, L.; Murphy, N.; Harte, E.; AH, M.; Hebestreit, H.; Dwek, R. A high-performance liquid chromatography based strategy for rapid, sensitive sequencing of N-linked oligosaccharide modifications to proteins in sodium dodecyl sulphate polyacrylamide electrophoresis gel bands. *Proteomics* **2001**, *1*, 285−294.
(47) Charlwood, J.; Skehel, J.; Camilleri, P. Immobilisation of antibodies in gels allows the improved release and identification of glycans. *Proteomics* **2001**, *1*, 275−284.
(48) Hirabayashi, J.; Arata, Y.; Kasai, K. Glycome project: concept, strategy and preliminary application to Caenorhabditis elegans. *Proteomics* **2001**, *1*, 295−303.
(49) Resing, K.; Mansour, S.; Hermann, A.; Johnson, R.; Candia, J.; Fukasawa, K.; Vande Woude, G.; Ahn, N. Determination of v-Mos-catalyzed phoshorylation sites and autophosphorylation sites on MAP Kinase Kinase by ESI/MS. *Biochemistry* **1995**, *34*, 2610−2620.

AR020143L