Methods and approaches for the comprehensive characterization and quantification of cellular proteomes using mass spectrometry

Shama P. Mirza and Michael Olivier

National Center for Proteomics Research, Biotechnology and Bioengineering Center, Medical College of Wisconsin, Milwaukee, Wisconsin

Mirza SP, Olivier M. Methods and approaches for the comprehensive characterization and quantification of cellular proteomes using mass spectrometry. Physiol Genomics 33: 3–11, 2008. First published December 27, 2007; doi:10.1152/physiolgenomics.00292.2007.—Proteomics has been proposed as one of the key technologies in the postgenomic era. So far, however, the comprehensive analysis of cellular proteomes has been a challenge because of the dynamic nature and complexity of the multitude of proteins in cells and tissues. Various approaches have been established for the analyses of proteins in a cell at a given state, and mass spectrometry (MS) has proven to be an efficient and versatile tool. MS-based proteomics approaches have significantly improved beyond the initial identification of proteins to comprehensive characterization and quantification of proteomes and their posttranslational modifications (PTMs). Despite these advances, there is still ongoing development of new technologies to profile and analyze cellular proteomes more completely and efficiently. In this review, we focus on MS-based techniques, describe basic approaches for MS-based profiling of cellular proteomes and analysis methods to identify proteins in complex mixtures, and discuss the different approaches for quantitative proteome analysis. Finally, we briefly discuss novel developments for the analysis of PTMs. Altered levels of PTM, sometimes in the absence of protein expression changes, are often linked to cellular responses and disease states, and the comprehensive analysis of cellular proteomes would not be complete without the identification and quantification of the extent of PTMs of proteins.

quantitative proteomics; isotopic labeling; phosphoproteomics

PROTEOMICS IS THE GLOBAL ANALYSIS of gene expression at the protein level. It has emerged as a postgenomic technology with the promise to unravel the cellular mechanisms of diseases and may lead to the development of reliable markers for disease diagnosis and therapy. Despite its promise, proteomic analyses remain challenging because of various limitations of existing technology. All commonly used approaches to isolate and characterize proteomes are still cumbersome.

Over the past several years, mass spectrometry (MS) has emerged as the most efficient and versatile tool of all the proteomics approaches available so far (1, 24, 26). With the emergence of electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) techniques, analyzing large molecules such as proteins by mass spectral analysis has become possible (31, 54, 112). Although significant progress has been made, current instrumentation technology does not allow the comprehensive and complete characterization of cellular proteomes and their posttranslational modifications (PTMs), and there is still a long way to go to reach this goal. Finishing the Human Genome Project was facilitated by the static nature of the genome, which does not change significantly over time within a normal cell. Conversely, the proteome, though encoded by the cellular genome, is not static and changes continuously through various cellular mechanisms. It is believed that the 25,000 genes in the human genome encode up to one million different proteins through various splice forms and posttranslational processing and modifications. Analyzing the genome is simpler because the DNA consists of just 4 building blocks, while proteins are composed of 20 different naturally occurring amino acids. To add to the complexity of the proteome, there is a large dynamic range of concentrations that changes over time from cell to cell and state to state.

Ideally, a comprehensive proteomic analysis needs efficient separation of the complex sample and subsequent identification, quantification, and complete characterization of proteins in a single experiment. Thus each comprehensive proteomic characterization involves a series of analytical techniques. Below, we describe and discuss commonly used approaches and technologies that help in the identification and characterization of proteins with MS.

Comprehensive Proteomic Characterization

A typical standard mass spectral analysis of a complex protein sample involves an initial fragmentation step, a separation methodology to fractionate the complex mixture, and a subsequent mass spectral analysis of the resulting fractions. Commonly, proteins in a complex mixture are fragmented with proteolytic enzymes such as trypsin. Upon treatment, a typical
sample from a whole cell lysate contains thousands to millions of peptides. These complex mixtures need to be fractionated before further analysis. Initial proteomic separation techniques were based on gel electrophoresis, either one-dimensional (1-DE) or two-dimensional electrophoresis (2-DE), in which proteins or peptides are separated based on their charge and molecular weight (19, 53, 69, 78). The separated proteins are usually visualized with different dyes. The 1-DE or 2-DE gel bands are excised and in-gel digested, and the resulting peptides are analyzed by either MALDI-MS or ESI-MS. Other methods to fractionate the peptide mixture include liquid chromatography and capillary electrophoresis (CE) (51, 86, 97, 129). While a successful protein and peptide separation and fractionation is essential to a mass spectral analysis, this review focuses on the analytical aspects and the challenges of MS for proteomics. Other reviews discuss the commonly used approaches and challenges of fractionation (see Ref. 69a).

MS is a powerful approach to obtain protein sequence data from unknown samples and to correlate the experimental data with sequence information in public databases (“bottom-up proteomics”; Fig. 1). Such protein or peptide sequence information can be obtained from tandem mass spectral analysis. In a typical tandem MS analysis, the first step consists of the detection of the initial peptide ion. Subsequently, the peptide ions are fragmented by collision-induced dissociation (CID) to break the polypeptide backbone at the amide bond, thereby creating a ladder of fragment ions that reflect the peptides’ amino acid sequence (11, 20, 118). The resulting spectra are then compared with publicly available protein sequence information. The observed masses of the proteolytic fragments are compared with theoretical in silico sequences to identify the peptide sequence and thereby the protein (14, 29, 36, 44, 65, 77). With the advancements in the separation techniques, up to 2,000 proteins can now be identified in a standard experiment.

Alternative approaches to this bottom-up approach, in which the original protein is predicted from sequence information of proteolytic peptides, have utilized “top-down” approaches in which the intact protein is directly dissected and amino acid sequence information is obtained by dissociation (Fig. 1). In this approach, the intact proteins are separated by gel electrophoresis or offline liquid chromatography before MS analysis. The major obstacle in this approach is the determination of product ion masses from multiply charged species of intact proteins. Because of the formation of multiply charged protein precursor ions, it is difficult to interpret top-down fragmentation spectra. This limitation can be evaded by reducing the charge states of the product ions through the introduction of gas-phase anions to strip protons from the product ions through ion-ion proton transfer reaction (106).

Another approach of choice for top-down proteomics is the use of the Fourier transform-ion cyclotron resonance (FT-ICR) mass spectrometer or orbitrap mass spectrometer with high mass accuracy (<2 ppm). The product ion charge state can be determined from the isotope spacing in the multiply charged species that facilitates the identification (27, 66, 85). The dissociation techniques in top-down proteomics are favored toward electron-capture dissociation (ECD) as implemented on FT-ICR-MS or electron-transfer dissociation (ETD) used in orbitrap instruments.

Although the top-down proteomics approach has some limitations because of the formation of very complex spectra, use of expensive instrumentation, and difficulties with proteins of high molecular mass (>50 kDa), it is advantageous over classical bottom-up approaches since it provides access to the complete protein sequence, locates PTMs due to gentle fragmentation methods, and avoids long protein digestion methods.

**Comprehensive Protein Quantification**

In addition to the identification of proteins, it is essential to quantify proteins in complex biological samples to better understand their role in an organism and in physiological systems (9, 21, 63, 68, 82, 100, 127). Both relative and absolute quantification methods have been established that use mass spectral techniques for the protein identification. Traditionally, protein quantification is carried out with gel-based approaches. In contrast, stable isotope labeling and label-free methods have been established that involve mass spectral analyses and result in quantification at the peptide level. Below, we discuss the different approaches in more detail and highlight the differences between the individual approaches.

**Gel-based quantification approaches.** Gel-based isolation of proteins was used extensively in molecular biology and biochemistry even before the development of modern MS-based protein identification. Even now it is used as a quantitative proteomic technology that can resolve hundreds to thousands of proteins in a single gel. To overcome the limitations of gel-to-gel reproducibility, differential gel electrophoresis (DIGE) has been established (83, 117, 122). In a
DIGE experiment, up to three different samples can be analyzed on a single gel, each labeled with a different fluorescent cyanine dye (Cy2, Cy3, and Cy5). Although the method is successful in quantifying proteins and identifying PTMs on proteins that result in lateral shift in spots on the gel, the approach has several limitations. Like any gel-based method, DIGE has limitations in resolving hydrophobic proteins, interference from high-abundance proteins, and poor resolution of spots. Furthermore, not all individual proteins that are differentially expressed can be subsequently identified with MS. Finally, multiple protein isoforms can often be found in different spots on the gel, complicating the comprehensive analysis.

Stable isotope labeling. With stable isotope labeling, both absolute and relative methods of quantification have been developed, each having its own merits and shortcomings. Most of the absolute methods of quantification are small-scale analyses targeted to specific proteins. In contrast, relative quantification methods are focused on large-scale global proteomic analyses. Any stable isotope labeling method involves labeling of proteins/peptides with one or more stable isotopes and pooling the sample with an unlabeled control sample. The mixed sample is then subjected to a mass spectral analysis. The signal intensities of the labeled and unlabeled peptides are used to measure the relative abundance, while further fragmentation of the peptide provides the sequence information necessary for identification of the quantified protein.

As illustrated in Fig. 2, stable isotope labeling can be classified into metabolic labeling, chemical mass tagging, and enzymatic labeling.

Metabolic isotope labeling requires in vivo incorporation of isotope-labeled essential amino acids during cell growth. The technology was first developed by Matthias Mann and his laboratory (1, 81, 82) and termed stable isotope labeling by amino acids in cell culture (SILAC). Cells are cultured in a medium supplemented with a labeled amino acid (lysine or arginine containing $^{13}$C or $^{15}$N). After several passages, cells are pooled with control cells grown in a medium of naturally occurring amino acids. The pooled samples are digested and analyzed by tandem MS. Although the technique is promising, it is limited to studies that involve cell culture so that cells can incorporate the exogenous amino acid into proteins. This normally precludes the use of this methodology for the analysis of tissue samples. However, a recent metabolic labeling study showed that entire model organisms (such as rats) can be labeled by using labeled chow. McClatchy et al. (71) demonstrated this quantification strategy by using $^{15}$N-labeled rat brain as an internal standard for large-scale analysis of the mammalian brain. The method is termed stable isotope labeling in mammals (SILAM) (71).

Mass tagging via chemical labeling is another approach using isotopes for quantification. In this method, proteins or peptides are tagged with a stable isotope-containing molecule.

---

**Fig. 2.** Illustration of commonly used labeling strategies for quantitative proteomic analysis using MS (adapted from Ref. 1). Proteins are depicted at top, the resulting labeled peptides from each approach are shown in the middle, and a schematic of the resulting mass spectrum is shown at bottom, with the resulting signal from the light peptides shown in gray and signal from the heavy peptides in red. Peak areas can be compared for relative quantification of peptide abundance.
Tagged peptides can be efficiently isolated and enriched with affinity groups attached to the tagging moiety that can bind to specific amino acid residues. One such technology is the isotope-coded affinity tag (ICAT) methodology developed by the Aebersold group (104). An affinity group is chemically linked to cysteiny1 residues on the protein. The ICAT reagent consists of a sulfhydryl-reactive iodoacetate group, a biotin affinity group, and a linker carrying light or heavy isotopes (101, 115, 123). After derivatization, both light and heavy tagged protein samples are pooled, digested, and enriched on an avidin affinity column to capture peptides containing tagged cysteine residues that can be analyzed by subsequent tandem mass spectral analysis. The method has both advantages and limitations due to the specificity of the affinity tag to specific amino acid residues. On one hand, the complexity of the sample will be reduced because of the enrichment of only tagged peptides. However, as a result, only a fraction of peptides will be analyzed by MS, and no quantitative or sequence information is obtained for proteins or peptides that do not contain cysteiny1 residues. While additional affinity tags have been developed in recent years that allow specific binding to other amino acid residues (34, 45, 61), the enrichment step is an essential part of the ICAT protocol and untagged proteins or peptides will not be analyzed and identified.

Recently, isobaric tags for relative and absolute quantification (iTRAQ) have been established as a more comprehensive and efficient method for proteomic quantification (2, 17, 94, 102, 128). The method was developed by Ross et al. (94). It involves tagging of the NH2-terminal and side-chain lysine amino groups with stable isotopically labeled mass tags. The reagent consists of an amine-specific reactive group, a mass balance group, and a reporter group. The reporter and balance groups carry stable isotopes, with different combinations of 12C, 14N, or 16O with their heavy isotopes 13C, 15N, or 18O, but since their molecular weights remain unchanged, the chromatographic properties of the peptides remain the same and the light and heavy peptides elute at the same time, enabling the quantification of different samples. After coelution, the peptides are further fragmented to release the iTRAQ reagent that allows the distinction of the different samples in MS and provides the necessary quantitative information. With the extensive success with four-plex reagents, eight-plex reagents are now commercially available that permit the quantification of eight different samples in a single run. The method can be used not only to perform relative quantification but also for absolute quantification by using an internal standard peptide labeled with one of the iTRAQ reagents. While the methodology has been used extensively, the iTRAQ system is relatively expensive, and the reporter ion masses are below the low-mass cutoff of ion-trap mass spectrometers, thus requiring more advanced mass spectrometers for the analysis.

Another nonisobaric triplex labeling strategy was introduced by Kellermann and Lottspeich (98, 99), named isotope-coded protein labeling technology (ICPL). In this approach, amino groups of intact proteins were derivatized with three isotopically different nicotinoyl reagents (12C6H4, 12C6D4, and 13C6H4). Since the method is based on labeling proteins even before tryptic digestion followed by fractionation, it significantly reduces the complexity of the sample but retains sequence coverage in the subsequent mass spectral analysis. This results in improved protein identification that is indispensable for the detection of PTMs.

Stable isotope incorporation by enzymatic labeling is another commonly used method to overcome the challenges of other methods discussed above. Enzyme-catalyzed labeling of peptides provides a comprehensive and global quantitation of the cellular proteome. Enzymatic incorporation of 18O atoms during proteolytic cleavage, most commonly by trypsin, results in peptides with either one or two 18O atoms at the carboxy terminus (8, 43, 74, 90, 108). The labeled sample is mixed with known amounts of an unlabeled peptide sample (for absolute quantification) or a reference sample (for relative quantification) and analyzed by tandem MS. In the 18O labeling strategy, the difference in molecular mass between light and heavy peptide samples is either 2 Da (one 18O atom incorporated) or 4 Da (two 18O atoms incorporated). The ratio of the intensities of the labeled and unlabeled peptides allows the quantification of the peptide. Although there are limitations in the method due to the back exchange of 18O with naturally occurring 16O, if sufficient care is taken (109), the method is advantageous over other labeling technologies because of its simplicity and global proteomic quantification both in vivo and in vitro samples. Our laboratory (41, 43) has developed protocols and analysis tools that allow the efficient quantitative analysis of complex biological samples using 18O labeling. Figure 3 illustrates the common work flow and the analytical software tool ZoomQuant.

**Alternative mass spectral quantification approaches.** Recently, label-free methods of quantification have been proposed (80, 116, 124). These methods are based on a correlation between peptide mass spectral peak data and the abundance of the protein in the sample. In one of the approaches, mass spectral peak intensities of peptide ions are used to quantify the protein amount (12, 16, 117). Alternatively, another approach involves spectral counting, where the number of mass spectra assigned to a protein is taken as a measure of protein abundance (33, 64). The label-free methods are advantageous, especially in samples where isotope labeling is either not possible or too cumbersome, and they provide quantitative information. However, the accuracy of the approach needs to be further evaluated, especially for low-abundance proteins in complex samples.

**Mass Spectral Analysis of Posttranslational Modifications**

The complex and dynamic nature of the cellular proteome has been a major challenge. The comprehensive assessment and analysis of the proteome is further complicated by various PTMs that regulate cellular processes (18, 56, 59, 72, 93, 114). Altered levels of PTMs, sometimes in the absence of protein expression changes, are often linked to cellular responses and disease states. Therefore, the comprehensive analysis of the cellular proteome would not be complete without the identification and quantification of the extent of PTMs of the individual proteins. Proteins are known to undergo more than 300 PTMs that regulate their cellular functions, of which phosphorylation is the most common and best characterized. Several
studies have described enrichment methods for these PTM proteins. However, even after enrichment, PTM proteins are not very well ionized under MS conditions because of their negative charge, complicating the MS detection of these proteins.

Classical approaches for phosphoproteomic analysis use polyacrylamide gel electrophoresis, as discussed above. The phosphorylated species can be visualized by radiolabeling, immunodetection, or phosphospecific staining (39, 67, 95, 120). The success in using gel-based methods is due to the isoelectric point (pI) shift of proteins with each phosphorylated site that shifts the protein spot on the gel laterally (28, 39, 49, 62, 119, 120, 125). However, all difficulties discussed above for gel-based quantification approaches apply to phosphoprotein quantification as well, and limit the use of the approach.

Mass spectral approaches to the analysis of phosphoproteins have been complicated by the technical difficulties of detecting phosphoproteins in the presence of nonphosphorylated species. Thus the phosphoproteins are commonly enriched before analysis. Various approaches have been established such as immunoprecipitation, chemical derivatization, affinity purification, and most commonly immobilized metal affinity chromatography (IMAC) methods (52, 91, 107). Below, we describe and discuss the individual approaches.

Immunoprecipitation. Immunoprecipitation of proteins from complex mixtures with antibodies against phosphorylated amino acid residues is a common approach in phosphoproteomics (25, 32, 47, 48). Most analyses have successfully used pTyr antibodies (84, 126), and recently attempts have been made to precipitate proteins with pSer and pThr antibodies (35). The precipitated proteins are further separated on 1-DE or 2-DE gels and analyzed by MS to map the phosphorylation site. While the method does allow the identification of phosphorylated proteins, and permits the characterization of the phosphorylated amino acid residue, the methodology does not allow the quantification of phosphorylated proteins.

Chemical derivatization. Chemical derivatization methods convert the phosphorylated amino acid into more traceable species. β-Elimination of pSer and pThr by alkali treatment is one of the most common derivatization approaches used (73, 92). β-Elimination of phosphoric acid from pSer or pThr followed by the introduction of a biotin moiety aids in the selective enrichment of phosphoproteins. The biotin-tagged phosphopeptides are selectively separated from nonphosphorylated species by affinity chromatography (34, 79). Another approach uses β-elimination followed by reaction with cysteamine to convert pSer or pThr to lysine analogs, which enables site-specific cleavage with trypsin at the phosphoamino acid site (55). However, the method has limitations due to the loss of chromatographic performance and loss of sensitivity during mass spectral analysis. Moreover, the method is only applicable to pSer and pThr residues.

Immobilized metal affinity chromatography. IMAC has been used extensively and efficiently for the enrichment of phosphoproteins and peptides (5, 15, 87). The approach exploits the high affinity of the phospho-moiety to positively charged metal ions like Fe³⁺ (5), Ga³⁺ (88), Al³⁺ (4), and Zr⁴⁺ (30). The metal ions are immobilized on a solid support (silica, Sepharose, or agarose) with metal-chelating agents such as iminodiacetic acid (IDA) (6), nitrilotriacetic acid (NTA) (3), tris(carboxymethyl)ethylendiamine (TED) (15), or poly(glycidylmethacrylate/divinylbenzene) (GMD) (6). Although the method has been used extensively, there are limitations to the approach. The acidic groups on the peptides are nonspecifically bound to the metal ions and hence need to be further esterified to block the acidic group, increasing the complexity of the experimental procedure (42). Moreover, the method is biased toward multiply phosphorylated peptides because of their high affinity to the metal ions. Finally, as with all described enrichment approaches, the method does not provide quantitative information.

Affinity purification. Affinity purification with metal oxides is an alternative method of phosphoprotein enrichment. Titanium oxide, aluminum oxide, and zirconium oxide have proven...
to be more chemically stable materials than silica-based materials. Recently, TiO₂ has been demonstrated (57, 96, 121) as an alternative reagent to the IMAC method of enrichment. Lately, Kweon and Hakansson (58) compared zirconium oxide to titanium oxide and concluded that each allows the selective enrichment of phosphopeptides, zirconium being more selective for monophosphorylated peptides while titanium had more affinity toward multiply phosphorylated peptides. Thus mixing both the metal oxides could increase the efficiency in the enrichment of both monophosphorylated and multiply phosphorylated species. Although the method is proven to be efficient in recovering over 90% of phosphopeptides, it has yet to be applied to complex protein mixtures.

An efficient MS-based method that not only identifies the PTM proteins but also measures the degree to which the protein is modified would be very informative for understanding protein function and signaling cascades in various cell functions. With the recently developed relative quantitation methods by MS, phosphoproteomic quantification has become possible (22, 37, 113). Recent studies have used SILAC, ICAT, iTRAQ, or 18O methods for the quantification of IMAC-enriched phosphopeptides followed by tandem mass spectral analysis (34, 89). Another approach is to dephosphorylate the peptides for the indirect identification and quantification of phosphopeptides (13, 103). The method is more advantageous over other methods because it only requires the MS analysis of unphosphorylated peptides, circumventing technical difficulties like signal suppression or biased enrichment procedures.

Despite the various enrichment methods for phosphopeptides, mass spectral analysis of these peptides is often not very efficient. The loss of the phospho-moiety from the peptides is very prominent during CID experiments. Thus a very intense peak with loss of 98 Da or to a minor extent at 80 Da dominates over the rest of the fragmentation peaks, thereby providing only limited data for peptide identification (10, 46, 50, 76). To overcome the limitations with CID analysis, new analytical methods like ECD (7, 40, 60, 105, 110) and ETD (23, 38, 70, 75, 111) have been established to identify phosphopeptides. The methods are based on gentle fragmentation of the peptides, preserving the labile amino acid modifications. In this technology, c and z ions are generated, keeping the modification site intact on the fragmented ions. Besides, there is no neutral loss of phosphoric acid from the peptide; even larger and multiply charged species are identified efficiently.

Perspectives

Mass spectral analysis of complex protein mixtures has been improved significantly over the past decade, and technological advances and improved methodologies now allow the comprehensive quantitative analysis of complex biological samples by MS. However, the methodologies are far from optimal, and only a small fraction of the entire proteome of a biological sample is analyzed quantitatively with any of the approaches described above. While further advances are needed to provide the data necessary for the complete dissection of proteomic changes in disease, current technology already allows an initial glimpse of the underlying pathophysiological changes on a cellular level. The information obtained from these initial proteomic analyses of carefully selected biological samples will significantly advance our current understanding of the complex regulation of biological systems, even with this rudimentary information. Analyses of temporal changes of cellular proteomes and the regulation and dynamic alteration of PTMs of the proteome will provide novel information beyond the genomic data currently being obtained. Once both approaches have been merged efficiently, we will be able to dissect the complex underlying pathways that regulate disease processes, and use this information for disease prevention and management.

GRANTS

S. P. Mirza and M. Olivier are supported by funding from the National Heart, Lung, and Blood Institute (N01-HV-28182).

REFERENCES


60.}
Review

Proteomics Using Mass Spectrometry


110. Sweet SM, Cooper HJ. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. J Proteome Res 5: 1214–1223, 2006.