Analysis of posttranslational modifications of proteins by tandem mass spectrometry

Martin R. Larsen, Morten B. Trelle, Tine E. Thingholm, and Ole N. Jensen

Protein activity and turnover is tightly and dynamically regulated in living cells. Whereas the three-dimensional protein structure is predominantly determined by the amino acid sequence, posttranslational modification (PTM) of proteins modulates their molecular function and the spatial-temporal distribution in cells and tissues. Most PTMs can be detected by protein and peptide analysis by mass spectrometry (MS), either as a mass increment or a mass deficit relative to the nascent unmodified protein. Tandem mass spectrometry (MS/MS) provides a series of analytical features that are highly useful for the characterization of modified proteins via amino acid sequencing and specific detection of posttranslationally modified amino acid residues. Large-scale, quantitative analysis of proteins by MS/MS is beginning to reveal novel patterns and functions of PTMs in cellular signaling networks and biomolecular structures.

INTRODUCTION

Determination of posttranslational modifications (PTM) of proteins is fundamental in elucidation of the intricate processes that govern cellular events, like cell division, growth, and differentiation. The term PTM denotes changes in the polypeptide chain as a result of either the addition or removal of distinct chemical moieties to amino acid residues, proteolytic processing of the protein termini, or the introduction of covalent cross-links between domains of the protein. PTMs are involved in most cellular processes including the maintenance of protein structure and integrity, regulation of metabolism and defense processes, and in cellular recognition events and morphology changes. Analysis of PTMs presents a number of challenges to protein and proteomics researchers, and efficient and sensitive methods for detection of PTMs are required. Traditionally, PTMs have been identified by Edman degradation, amino acid analysis, isotopic labeling, or immunoochemistry. Within recent years, mass spectrometry (MS) has proven to be extremely useful in PTM discovery. The presence of covalent modifications in proteins affects the molecular weight of the modified amino acids, and the mass increment or deficit can be detected by MS (Table 1). MS has several advantages for characterization of PTMs, including (i) very high sensitivity; (ii) ability to identify the site of PTM; (iii) discovery of novel PTMs; (iv) capability to identify PTMs in complex mixtures of proteins; and finally (v) the ability to quantify the relative changes in PTM occupancy at distinct sites. None of the other techniques provide all these features. In this review we describe the utility of tandem mass spectrometry (MS/MS) for the determination of PTMs and provide selected examples of recent modification-specific proteomic studies.

MS IN PROTEOMICS

MS is now widely used in protein biochemistry and in proteomics for the identification and characterization of proteins in cell lysates, isolated organelles, or purified multisubunit complexes (1–3). Protein separation technologies based on centrifugation, electrophoresis, or chromatographic methods are readily interfaced to MS in an online or off-line fashion. Proteins are then converted into peptides by treatment with sequence-specific proteases or chemical reagents, since peptides are more amenable to MS and MS/MS analysis than intact proteins.

The availability of robust and sensitive matrix-assisted laser desorption/ionization MS (MALDI-MS) (4) and electrospray ionization MS (ESI-MS) (5) instruments makes advanced MS technology accessible to molecular cell biologists, biochemists, and proteomics researchers. MS using either MALDI or ESI as the ionization method enables accurate mass determination of peptides. Peptide mass fingerprinting by MALDI-MS and subsequent sequence database searching is widely used for identification of proteins that are isolated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) or by two-dimensional gel electrophoresis (2DE) gels. In most studies it is also desirable to obtain amino acid sequence information, because the accurately determined molecular mass in combination with a partial amino acid sequence of a peptide are very specific probes for protein identification (1,6). Furthermore, mass
analysis and amino acid sequencing by MS/MS may reveal the presence of PTMs at individual amino acid residues in proteins (3, 7).

PEPTIDE SEQUENCING BY MS/MS

The tandem mass spectrometer provides the means for amino acid sequencing of peptides. It enables gas-phase isolation of individual peptide ion species followed by collision-induced dissociation (CID) and detection of the resultant amino acid sequence specific fragment ions (Figure 1). The MS/MS experiment consists of several stages of mass analysis and ion manipulation. First, the masses of the sample analytes are determined in a survey scan (first MS experiment). Second, the peptide ion of interest is isolated via its mass-to-charge ratio (m/z) value (i.e., by filtering away other ion species that have a different m/z value). Third, the selected peptide ion species is activated (e.g., by collisions with an inert gas such as Argon that imparts internal energy into the ions and thereby induces their fragmentation). Last, the m/z values of the fragment ions are determined (second MS experiment). The most labile bonds in peptides are generally the backbone amide bonds, leading to breakage of the peptide backbone in between amino acids. Hence, tandem mass spectra of peptides contain a series of sequence-revealing fragment ions (8). The fragment ion signals reflect the amino acid sequence as read from either the N-terminal (b-ion series) or the C-terminal (y-ion series) direction (9). The identities of the individual amino acids are revealed by the mass differences between the signals in b-ion series or y-ion series (Figure 1).

MS/MS is a sensitive and fast technique for peptide sequencing. Depending on the type of instrumentation and the experimental design, 10–500 ng pure protein are needed to obtain sufficient peptide data for protein identification. Micromgram amounts of protein are needed for proteomics analysis of complex protein samples and organelles. Coupling of capillary liquid chromatography to tandem mass spectrometry (LC-MS/MS) allows for automated analysis of complex peptide mixtures by data-dependent acquisition of peptide mass spectra and tandem mass spectra. Typically, 1–5 peptides are analyzed by MS/MS per second. Hence, hundreds or even thousands of peptides can be analyzed in a single LC-MS/MS experiment to reveal the inventory of proteins in biological samples.

Large-scale MS experiments in proteomics are facilitated by computational data analysis and database searching algorithms that enable protein identification and quantitation. A series of computational methods for the assignment of peptide sequences based on automated interpretation of MS/MS spectra and protein sequence database searching have been developed (6, 10, 11). These tools are often interfaced to or integrated with MS software provided by manufacturers to give an efficient data processing pipeline for annotation of peptides and the equivalent proteins.
PTM ANALYSIS BY MS/MS

Accurate determination of the mass increment at the protein or peptide level will aid in defining the type of modification. Intact mass determination of purified proteins is a useful method for determining modification and processing events. Comparison of the experimentally detected intact molecular mass with the calculated mass obtained from the amino acid sequence of the protein will reveal any discrepancies, and the mass difference or deviation may define the modification. In addition, the mass spectrum will often reveal heterogeneous modification and processing of the protein, resulting in multiple species that each correspond to an individual modification state.

A large class of PTMs is represented by chemical moieties that are covalently attached to proteins by various enzymes. Examples are phosphorylation (+80 Da), sulfation (+80 Da), nitration (+45 Da), O-glycosylation (>203 Da), and acylation (>200 Da) (Table 1 and references therein).

Another class of PTM is proteolytic processing that leads to mass deficit. Removal of the leader methionine residue (131 Da) is an example of such a processing event. Processing of precursor polypeptides for removal of signal sequences to generate biologically active proteins is yet other types of processing. Amino acid oxidations are also common PTMs, especially on methionine and tryptophan residues, leading to mass increments of 16 or 32 Da (Table 1). The formation of disulfide bridges between cysteine residues leads to a mass change of -2 Da. Thus, mass analysis of intact proteins or the derived peptides can reveal modifications and thereby also the detection of site-specific PTMs (3).

The chemical stability of the PTM is decisive for its efficient detection in MS/MS. Certain PTMs will remain intact during MS and MS/MS experiments. For example, acetyl-lysine is a very stable PTM that leads to a mass increment of 42 Da of the intact peptide. Upon MS/MS of a lysine-acetylated peptide, all the fragment ions that contain the AcLys residue exhibit a +42 Da mass increment relative to the unmodified peptide (Table 1). Less stable PTMs are phosphoserine and phosphothreonine, which often eliminate phosphoric acid (or a phosphate group and water) during MS/MS (12).

In addition to the amino acid sequence-specific peptide fragment ions, also PTM-specific signals from modified amino acids exist. These PTM-specific fragment ions appear in the lower mass range of tandem mass spectra, typically below 1000 Da. This makes them ideal for the detection of site-specific PTMs.

Table 1. Mass Values for Some Posttranslational Modifications and Frequently Observed Neutral Losses and Diagnostic Ion Signals

<table>
<thead>
<tr>
<th>PTM</th>
<th>Δm (MS) (Da)</th>
<th>Δm (MS/MS)</th>
<th>Neutral loss (Da)</th>
<th>MS/MS Diagnostic ion (m/z)</th>
<th>References (examples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphotyrosine</td>
<td>79.966</td>
<td>79.966</td>
<td>-78.959 (neg. mode) 216.0422 (pos. mode)</td>
<td>16,17</td>
<td></td>
</tr>
<tr>
<td>Phosphothreonine</td>
<td>79.966</td>
<td>97.977</td>
<td>-78.959 (neg. mode)</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Phosphoserine</td>
<td>14.016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimethyllysine</td>
<td>28.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylarginine</td>
<td>14.016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimethylarginine</td>
<td>28.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trimethyllysine</td>
<td>42.047</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-linked glycans</td>
<td>162.053</td>
<td>162.053</td>
<td>163.061</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>203.079</td>
<td>203.079</td>
<td>204.087</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>291.095</td>
<td>291.095</td>
<td>292.103, 274.093</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>365.148</td>
<td>365.148</td>
<td>366.156</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(HexHexNAc)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitrotyrosine</td>
<td>44.985</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palmitoyl</td>
<td>238.230</td>
<td>272.217</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acylation</td>
<td>Various, see ref</td>
<td>See ref</td>
<td>See ref</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>114.043</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methionine oxidation</td>
<td>15.995</td>
<td>63.998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tryptophan oxidation</td>
<td>3.995, 15.995, 31.990</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PTM, posttranslational modification; MS, mass spectrometry; MS/MS, tandem mass spectrometry; m/z, mass-to-charge ratio.
MS/MS provides a number of useful analytical features that take advantage of diagnostic ion signals and neutral loss products. Precursor ion scanning is a selective and sensitive method for specific detection of only those peptides that carry a PTM that generate a unique diagnostic ion. For example, phosphopeptides generate a strong m/z 79 ion signal in the negative ion MS/MS mode. Thus, precursor ion scans for this ion will identify only phosphopeptides in the sample, whereas regular peptides remain undetected. A number of glycans and lipids generate fragment ions that can be explored for specific detection of PTM peptides (Table 1). The precursor ion scanning method is extremely useful for the analysis of phosphorylation sites and other PTMs in individual, purified proteins (12,17–23) (Table 1).

Neutral loss analysis is in proteomics typically performed by ion trap instruments that allow very fast MS/MS and multistage mass spectrometry (MS/MS/MS) analysis (24–26). The principle is based on the lability of PTMs, such as phosphoserine/threonine, that readily undergoes gas-phase elimination of a PTM-related chemical group during MS/MS. For example, the neutral loss of H₃PO₄ observed upon MS/MS of phosphopeptides will trigger a MS/MS/MS analysis of the product, to reveal the sequence of the PTM peptide and the position of the PTM (Figure 2).

### COMPUTATIONAL TOOLS FOR PTM ASSIGNMENT IN MS/MS DATA SETS

Several computational methods for automated annotation of PTM in peptides exist. These algorithms analyze the MS and MS/MS data, taking into account the Δm values and sometimes also neutral losses and diagnostic ions for the PTM of interest. Once a set of proteins is identified in a MS/MS-based proteomics experiment, it is possible to perform a more detailed analysis of the retrieved protein sequences and the MS/MS data set. The initial database search is typically performed with only a limited set of specified modifications, including alkylated cysteines and oxidized methionines, in order to minimize search space and to avoid false positive PTM assignments. In a second stage, the molecular masses of the unmodified proteins and the corresponding proteolytic peptides are readily calculated from the protein sequences obtained by the initial sequence database search. As mentioned previously, PTM will lead to a mass increment or mass deficit (Δm) of the modified peptide relative to the unmodified species. Thus, it becomes feasible to inquire whether any of the initially unassigned MS/MS spectra can be matched to posttranslationally modified peptides, given a list of putative modifications and their Δm values and the list of predicted and observed peptide masses.

Although highly useful, computational sequence annotation tools should be used with care, because they may produce false positives in cases where the mass accuracy or signal-to-background level of the MS and MS/MS data are not sufficient to
make unambiguous assignments. This is particularly important in modification-specific proteomics, in which protein identification is often based on the detection and sequencing of only a few PTM peptides per protein. The presence of multiple modifications or various different types of modifications in a peptide may also complicate MS and MS/MS data interpretation.

STRATEGIES FOR DETERMINATION OF PTM IN PROTEOMICS

Mapping of PTMs in proteomics is a demanding task because most PTMs are low abundance and/or substoichiometric and some PTMs are labile during MS and MS/MS. In addition, many modifications are hydrophilic, which complicates PTM sample handling and purification prior to MS (27). The presence of PTMs may affect the cleavage efficiency of proteases, such as trypsin, to generate unexpected or large peptide products. Certain PTMs will reduce the ionization and detection efficiency in MS. Multisite PTMs may generate very complicated MS and MS/MS data sets that are difficult to interpret. For these reasons, it is often useful to consider and explore several approaches for mapping of PTMs in proteomics (Figure 3).

Generally, it is recommended to reduce the complexity of the sample as much as possible prior to MS analysis for mapping of PTMs. It is not advisable to pursue PTM mapping by direct LC-MS and MS/MS analysis of crude cell lysates, as PTMs are rarely detected in such experiments. This is mainly due to the presence of many abundant unmodified peptides, the inadequate peptide separation by LC, and the limitations of MS/MS for peptide sequencing of very complex mixtures. Biochemical purification of cellular compartments, organelles, protein complexes, or of individual proteins has proved to be a successful first stage for analysis of PTMs, as it dramatically reduces the complexity of the initial protein sample. As an alternative, PTM-specific reagents are useful for selective enrichment of PTM proteins and peptides prior to mass spectrometry.

Mapping of PTMs in proteins is achievable by a series of approaches, including (i) treatment with multiple proteases and shotgun sequencing by MS/MS; (ii) enrichment of modified proteins or peptides prior to MS/MS sequencing; (iii) and PTM-specific MS/MS and multistage MS. Combinations of these methods have proven successful for comprehensive analysis of PTM in proteomics (Figure 3).

Shotgun Sequencing

The shotgun sequencing approach is based on using multiple proteases with different cleavage specificity to generate complementary and redundant sets of overlapping peptides. The peptide mixtures are then analyzed by LC-MS/MS or by multidimensional separation technology and MS/MS (LC/LC-MS/MS) for comprehensive sequencing of proteins. Database searching and computational assembly of the sequence data then provides an overview of the protein sequences and annotation of PTMs and amino acid substitutions (28,29), and it is also amendable to mapping of modifications in purified proteins and protein complexes.

PTM-Specific Enrichment

Large-scale analysis of PTMs is facilitated by PTM-specific protein and peptide enrichment methods, such as PTM-directed affinity chromatography or immunoprecipitation with PTM-specific antibodies. Phosphoproteins can be purified by PTM-specific affinity resins (30) or anti-phosphoamino acid antibodies (31,32). Subsequent digestion of protein and LC-MS/MS analysis of peptides will reveal the identity of the retrieved proteins and sometimes allow site-specific assignment of phosphorylation sites in the recovered proteins. However, it is often also beneficial to enrich for PTM peptides prior to MS analysis in order to improve sensitivity and specificity. Phosphopeptides can be recovered by antibodies (33), immobilized metal affinity chromatography (IMAC)
(34–36, or by TiO2 columns (37,38) prior to MS/MS analysis. Strong cation exchange and anion exchange chromatography have also proven useful for reducing peptide complexity prior to MS/MS based detection and sequencing of modified peptides, including phosphopeptides (24,39). Using combinations of these methods, large-scale phosphoproteomics have revealed thousands of phosphorylated sites in proteins from various species (24,26,39,40).

Glycoproteins can be enriched by using lectins, which are sometimes referred to as sugar-specific antibodies (41,42). Glycoproteins from serum were retrieved and identified by a two-stage lectin enrichment procedure that first targeted the intact glycoproteins and then the proteolytically derived glycopeptides. The glycopeptide sample was then treated with N-glycosidase F in the presence of 18-O water and analyzed by LC-MS/MS. The 18-O labeling enabled site-specific assignment of glycosylation sites by signature ion signals MS/MS (43). An alternative strategy used lectins for glycoprotein enrichment followed by hydrophilic interaction chromatography (HILIC) for glycopeptide enrichment. Glycosidase D/H treatment and MS/MS then facilitated protein identification and assignment of glycosylation sites (44). GPI-anchored proteins, another class of glycoproteins, were identified by using modification-specific enzymes (phospholipases) to selectively release GPI-anchored proteins from plasma membrane preparations in a two phase detergent system, followed by their identification by LC-MS/MS and bioinformatics sequence analysis (45). O-glycosylation is also becoming tractable by using affinity enrichment and advanced MS methods (46).

Genetic methods are also useful for enrichment of PTM proteins. Ubiquitylated and SUMOylated proteins are amendable to purification and identification by integration of genetic tags into the modifying proteins (47).

PTM-Specific Chemistry

Certain modifications are amenable to chemical conversion to stable, tractable species. This is particularly useful for PTMs that are labile during MS and MS/MS analysis. O-phospho-Ser/Thr and O-GlcNAc-Ser/Thr readily undergo β-elimination to generate intermediates that are substrates for Michael addition of alkylating groups. Such methods have been used for selective recovery of phosphoproteins and glycoproteins by solid-phase chemistry or by affinity purification prior to protein digestion and identification by MS/MS sequencing (46,48). The introduction of fluorous affinity tags helps recover modified peptides by taking advantage of the unique chromatographic properties of fluorous compounds (19). Another option is to use β-elimination/Michael addition reactions to introduce unique mass tags into the modified peptides. The chemical design of the mass tag then provides a diagnostic mass fingerprint for the modified peptides, so they are readily distinguished from regular peptides (49). Other chemistries and resins, including dendrimers, have also been applied for PTM-specific recovery of proteins and peptides (50–52). Tagging-by-substrate methods were applied to identify O-GlcNAc-modified proteins and acylated proteins (53,54).

QUANTIFICATION OF PTM

Dynamic cellular events, such as signal transduction networks, cell cycle progression, and chromatin activity call for quantitative techniques for determination of PTMs. MS-based methods facilitate both absolute and relative quantitation of peptides and their PTM. Absolute quantitation of peptides by MS is achievable by using internal standard peptides for selected proteins and defined PTMs (55). Relative quantitation of PTM peptides is obtained by either peptide intensity profiling (PIP) by LC-MS or by stable isotope labeling (SIL) by using stable isotope-encoded chemical precursor molecules or alkylating reagents (1,3). The functionally important PTMs will exhibit variations in abundance as a result of the perturbation, and they are identified by comparison of the peptide ion signals obtained from two or more defined cellular states, typically a control experiment and one or more perturbed states. The relative abundance of individual peptides is then derived by comparative analysis of LC-MS data using retention time and mass (PIP) or by using only the MS ion intensity of isotopically encoded peptides (SIL). These methods have been used in quantitative phosphoproteomics in various organisms, from microbes to humans (26,40,56–58), and also in glycoproteomics (43,59,60).

CONCLUSIONS AND PERSPECTIVES

MS/MS is increasingly integrated in modern cell biology and biomedical research, in academia and in industry. Qualitative and quantitative MS-based strategies for PTM mapping in proteins have already contributed detailed insights into a series of complex biological systems, from microbes to plants and mammals. Novel MS technologies are continuously developed. Recent examples include the introduction of robust high-performance hybrid tandem mass spectrometers that provide very high mass accuracy [mass error <5 parts per million (ppm)] and multistage MS capabilities at a chromatographic timescale, thereby allowing detailed analysis of PTMs during LC-MS. Novel and improved MS/MS ion dissociation methods, such as infrared multiphoton dissociation (IRMPD) (61), electron capture dissociation (ECD) (62), and electron transfer dissociation (ETD) (63) have already proven useful for mapping of PTMs in biological systems. Increasingly advanced computational methods are useful for prediction of PTMs prior to MS analysis of known proteins, as well as for the identification of PTM peptides in large-scale data sets. Multistage MS and efficient fragmentation methods such as ECD and ETD also facilitate top-down analysis of intact proteins and large polypeptides (64). The analytical potential of these novel methods for mapping of PTMs in proteomics research is currently explored in many laboratories worldwide. No doubt, the prospects for MS-based analysis of PTM of protein are very exciting.
ACKNOWLEDGMENTS

We thank our colleagues in the Protein Research Group for their contributions and for discussions. This work was supported by grants from the Danish Natural Sciences Research Council and the Strategic Research Agency (Young Investigator Awards to M.R.L. and O.N.J.). O.N.J. is a Lundbeck Foundation Research Professor.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

Elortza, F.M. White.


B.


261

46. Elortza, F.M. White.


44. Elortza, F.M. White.


34. Elortza, F.M. White.


2. Elortza, F.M. White.