



Current Trends in the Analysis of Post-translational Modifications

Dávid Virág¹ · Borbála Dalmadi-Kiss¹ · Károly Vékey² · László Drahos² · Imre Klebovich¹ · István Antal¹ · Krisztina Ludányi¹

Received: 29 April 2019 / Revised: 23 July 2019 / Accepted: 19 August 2019 / Published online: 31 August 2019
© The Author(s) 2019

Abstract

Post-translational modifications controlling a large number of biological functions are key aspects of protein diversity. They have an important role controlling cellular processes and may be advantageously utilized. Qualitative and quantitative analyses of post-translational modifications are useful for biomarker research and an integral part of the characterization of protein biopharmaceuticals. Due to its sensitivity and widespread applicability, mass spectrometry has become the core technology of the analysis especially when combined with chromatographic and other separation techniques. The aim of this article is to present a general overview of mass spectrometry applications in the field of PTM mapping. We also present the analytical challenges of particular PTMs, primarily focusing on the most frequent modifications.

Keywords Protein · Post-translational modification · Mass spectrometry · Liquid chromatography

Introduction

One of the main goals of molecular biology is to understand the working mechanism of cells or even the whole organism. In this context, an essential step is to identify and characterize processes present in cells at the molecular level. An important milestone was the Human Genome Project, in which the sequence of 3 billion nucleotide base pairs that build up human DNA pool were identified, and all the genes have been mapped [1]. The genome is almost identical in every human cell and is nearly constant during the lifespan of an organism. In contrast, the proteins in a cell are much more numerous, vary significantly both among various cells and during the lifetime of a cell [2]. While the genome contains between 20,000 and 25,000 genes, the transcriptome consists of approximately 100,000 transcripts and the proteome exceeds 1 million proteins (Fig. 1) [3]. The expansion from the genome to the proteome is based on

two major mechanisms during transcription and protein synthesis. Most eukaryotic genes contain multiple non-coding intron sequences, which are copied into the precursor messenger RNA (pre-mRNA). During the mechanism of splicing, these sequences are removed from the pre-mRNA and exons are joined together, forming the mature mRNA. Exons can be ligated together in multiple ways, providing different mRNAs and protein isoforms. This phenomenon—so-called alternative splicing—is one of the key generators of protein diversity [4]. The second way to increase complexity is the covalent and commonly enzymatic alteration of the folded or the nascent protein during or after protein biosynthesis. These so-called post-translational modifications have two major categories: covalent addition of a functional group and cleavage of the peptide bond in the process of converting a propeptide to the mature form. However, in a broader context, non-enzymatic addition of a modifying group, autocatalytic cleavages and other structural changes (e.g. formation of disulfide bridges, cyclisation of asparagine or aspartic acid residues, racemization, etc.) can be considered as post-translational modifications as well [5].

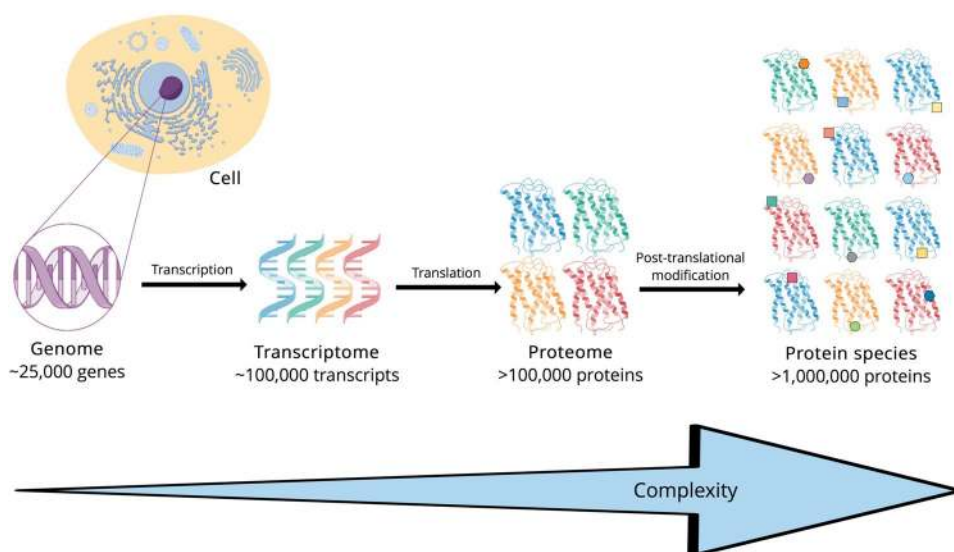
A vast number of chemical groups can attach as PTM to different parts of the peptides and proteins, greatly increasing the chemical repertoire of the 20 standard amino acids. The addition of modifying groups may take place on electron-rich functional groups that act as a nucleophile in the process, such as hydroxy groups of tyrosine, threonine,

✉ Krisztina Ludányi
ludanyi.krisztina@pharma.semmelweis-univ.hu

¹ Department of Pharmaceutics, Semmelweis University, Hőgyes Endre utca 7, Budapest 1092, Hungary

² MS Proteomics Research Group, Institute of Organic Chemistry, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, Magyar tudósok körútja 2, Budapest 1117, Hungary

Fig. 1 The increase in complexity from the genome to the proteome



serine, amino groups of arginine and lysine, carboxyl group of glutamate and aspartate or at the C or N termini of the protein [6]. Table 1 lists the most frequent types of PTMs.

The biological function of over 200 PTMs have been reported, including cellular localization, turnover and activity state of proteins, modulation of gene expression, activation of enzymes, interactions, etc. [6]. Nevertheless, the physiological relevance of many PTMs is still unknown, or only partially understood. The significance of post-translational modifications goes far beyond their role in cellular, physiological processes. A vast number of analytical data indicate that in various physiological and/or pathological states (pregnancy, inflammation, cancer) the structure, sugar composition, and the absolute and relative amount of various glycoforms change significantly, and often anomalous molecule variants can also be detected. It seems that various pathophysiological processes are reflected by the qualitative and

quantitative fingerprints of the antennary oligosaccharide structure of glycoproteins [12–19]. A number of ongoing research projects pay particular attention to the altered glycosylation and diagnostic value of circulating glycoproteins and use them as promising biomarkers. Altered phosphorylation and acetylation are also associated with some diseases [20, 21].

The biopharmaceutical industry has been growing constantly over the past few decades. Today, hundreds of therapeutic proteins are in medical use, the majority of which—according to the type of expression system—bear simple or complex PTMs, highlighting the importance of such structural features. Their significance lies in their ability to influence functional activity, stability, and immunogenicity of the product. Detailed characterization of PTMs is, therefore, indispensable during the development of both original biopharmaceutical products and biosimilars [22].

Table 1 Most frequent post-translational modifications, their localization, and functions [8–11]

PTM type	Frequency ^a	Localization	Main functions
Phosphorylation	58,383	Serine, threonine	Regulation of enzyme activity
Acetylation	6751	Lysine, N terminus	Subcellular localization, protein stability, regulation of DNA–protein interactions
O/N-linked glycosylation	6659	Serine, threonine/asparagine	Protein folding/stability, cellular adhesion
Amidation	2844	C terminus	Signal transduction, receptor recognition
Hydroxylation	1619	Proline, lysine	Stability of collagen
Methylation	1523	Lysine, arginine	Regulation of protein–protein interactions
Ubiquitination	878	Lysine, cysteine, serine, threonine, N terminus	Proteosomal degradation
Pyroglutamic acid	826	Glutamic acid, glutamine	Protein stability
Sulfation	504	Tyrosine	Strengthening protein–protein interactions

^aFrequency refers to the relative abundances of each PTM found putatively and experimentally in Swiss-Prot database [7]

Despite the great significance of PTMs, their large-scale investigation has been hindered for a long time by the absence of suitable analytical techniques. Mass spectrometry, especially coupled to high performance liquid chromatography, has often become the method of choice for qualitative and quantitative PTM analyses [23]. In the present article, we give an overview of the role of mass spectrometry for analysis of post-translational modifications, with particular emphasis on the most common ones, such as phosphorylation, acetylation, O/N glycosylation, amidation, and hydroxylation.

Principles and Trends in Mass Spectrometry-Based PTM Mapping

Analysis of proteins requires soft ionization techniques, such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI), that are able to ionize thermolabile and non-volatile compounds without considerable fragmentation [24]. In MALDI, samples are co-crystallized with a large amount of UV-absorbing matrix material on a metal plate and are bombarded by a pulsed UV laser, generating volatilized and typically +1 charged particles [25]. However, novel matrices support the formation of multiply charged ions providing improved fragmentation efficiency and enabling the use of electron-based fragmentation methods, which can be highly beneficial in the analysis of labile PTMs (see later in the section) [26]. This ionization technique allows direct analysis of mixtures and biomolecules in a wide molecular mass range (~300–200,000 Da). Additionally, it is considerably tolerant of salt contamination [25]. The most important disadvantage of MALDI is that it is not straightforward to couple it to liquid chromatography on-line. Furthermore, quantitation using MALDI is often considered inaccurate. In ESI, a liquid sample solution passes through an electrically charged capillary to the ion source, where a nebulizer gas helps ionization. As there is a few kV potential difference between the electrospray tip (capillary) and the entrance of the mass spectrometer, and the liquid leaving the capillary forms charged liquid droplets [27], multiply charged (protonated or cationized) ions are generated from these droplets. This makes ESI well suited for analysis of both small and large molecules, such as peptides, proteins, glycoproteins, oligonucleotides. Electrospray has better reproducibility than MALDI and is far better suited for quantitation. Direct coupling of ESI with separation techniques (such as HPLC or CE) is typical, and quite straightforward.

Structure analysis by mass spectrometry, and especially by tandem mass spectrometry, is based on fragmentation (the nascent molecular ion producing fragment ions by specific chemical reactions). Excitation of the ion to be studied

(often, but not always the molecular ion) is accomplished by a variety of methods, most typically collision-induced dissociation (CID) is used (sometimes, also called collisionally activated dissociation, CAD). In this process, the precursor ion (accelerated in the mass spectrometer) collides with a neutral gas in the collision chamber. In the case of peptides and proteins, this results in a cleavage of the peptide backbone, yielding mostly b- and y-type fragments [28], which can be used to determine the peptide sequence. When the peptide bears labile modifications (e.g. phosphorylation, sulfonation), the predominant cleavage site may be redirected, often leading to loss of the labile modification. As a consequence, the spectra may be inadequate to determine the peptide sequence and/or the localization of modification.

Electron capture dissociation (ECD) is an alternative of CID. In ECD, the precursor ion (e.g. a multiply protonated peptide) captures one (or more) low-energy electrons, and then quickly dissociates. ECD generates cleavage at N-C α bond of the peptide backbone, providing c- and z-type fragment ions. Probably, the most engaging characteristic of this fragmentation method is that labile PTMs generally remain intact. An alternative of ECD is electron transfer dissociation (ETD) [29], showing similar fragmentation.

Detailed characterization of the PTM of a protein (or set for proteins) is a challenging task. First of all, these modifications are generally heterogeneous, thus the amount of a given PTM variant (isoform or glycoform) is only a fraction of the total protein amount. Second, the dynamic range of human proteome (i.e. the concentration of abundant or minor proteins/isoforms) exceeds several orders of magnitude, aggravating identification and quantification of low-abundance proteins or isoforms. Third, full structural characterization of a protein requires much more substance and analysis time than a “simple” identification based on a few peptide fragments [23]. Fourth, biological samples are often available in limited amounts only, such as tumor biopsies or tissue slices [30]. These complications highlight the importance using elaborate sample pretreatment, separation and enrichment methods for protein analysis. A large number of strategies have been developed to recover sufficient amount of proteins even if the amount of biological material is very small (< 100 μ g) [30]. Although development and implementation of such techniques is time-consuming, they are indispensable to decrease the complexity of the biological sample prior to MS analysis.

The commonly used methods can be divided into gel-based and gel-free approaches [31]. Generally, the former works better at the protein level, the latter at the peptide level. Gel-based techniques usually employ sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) or two-dimensional gel electrophoresis (2-DE), in which the first dimension refers to separation based on the pI, then the analytes are separated according to their molecular

weight in the second dimension. Various procedures, for instance, immobilized pH gradient, can improve sensitivity and selectivity to get a more exhaustive picture of the analytes. Despite its favorable efficiency, the method has a number of limitations, such as slowness and incapability for high throughput analysis. The use of gel electrophoresis is further aggravated by high dynamic range of proteins in biological samples. Gel-based techniques are often followed by MALDI analysis. In such case, the protein line or spot is cut out, digested in the gel and sampled by MALDI [32], or (in more complex cases) may be analyzed by (nano)LC–MS/MS [30]. Gel-free techniques include a variety of liquid chromatographic techniques and capillary electrophoresis (CE) [33]. Amongst several multidimensional LC methods, the most commonly applied technique separates the peptide mixture using strong cation-exchange (SCX) chromatography, followed by RP-LC–MS/MS. The principle behind this is that peptides are first separated based on their charge, and in the sequel based on their hydrophobicity. SCX chromatography can be coupled on-line as well as off-line to RP-HPLC [33].

Mass spectrometric approaches for PTM analysis and for “conventional” protein identification are analogous, and may be considered bottom-up, middle-down, and top-down proteomics [34]. In the bottom-up approach, peptides (typically in the 500–3000 mass range), produced by the enzymatic digestion of proteins, are studied. When this is performed on a large set of proteins, such as plasma or cell lysate, this is often called “shotgun proteomics” [35]. This approach is the most widespread and popular, owing to its sensitivity and high throughput [36]. As it has mentioned above, bottom-up approach works at the peptide level. A serine protease, trypsin is generally employed for this purpose, which cleaves the peptide bonds at the carboxyl side of basic amino acids, such as arginine and lysine generating oligopeptides in the optimal mass range for chromatographic separation, and the emerging peptides can be ionized properly in ESI due to their basic amino acid content, yielding satisfactory information to determine the parent protein. On the other hand, this strategy has serious limitations, which make bottom-up approach suboptimal for PTM mapping and identification of alternatively spliced protein variants. The main reason is that only a fraction of the peptide fragments is usually observed in MS analysis, resulting in a sequence typically in 20–60% range (and there will be no information of the rest of the protein). Tryptic digestion may also result in very short peptides, that are inappropriate for LC–MS analysis, and even if observed, do not contain much information. This is often a problem for analyzing Arg- and Lys-rich proteins, such as histones [37].

In the case of top-down proteomics, intact proteins are directly introduced and fragmented in a tandem mass spectrometer, without being subjected to proteolytic digestion.

Middle-down proteomics is an in-between case, which can be considered as a mixture of top-down and bottom-up methods. This procedure works with 5–20 kDa large polypeptides, produced by limited proteolytic digestion [34]. Figure 2 presents workflows with respect to these proteomic strategies.

Working at the protein level (i.e. with the top-down approach) offers a number of advantages in the field of PTM mapping. Since the whole protein is studied, there is no complication due to the different ionization efficiencies of peptide fragments. The top-down approach is able to identify sequence variants and PTM combinations as well [38]. It is particularly useful for the characterization of small proteins [38]. On the other hand, top-down proteomics is generally restricted to individual proteins or simple mixtures [39]; has relatively low sensitivity and data evaluation is very time-consuming; therefore, it is unsuitable for high-throughput work [40].

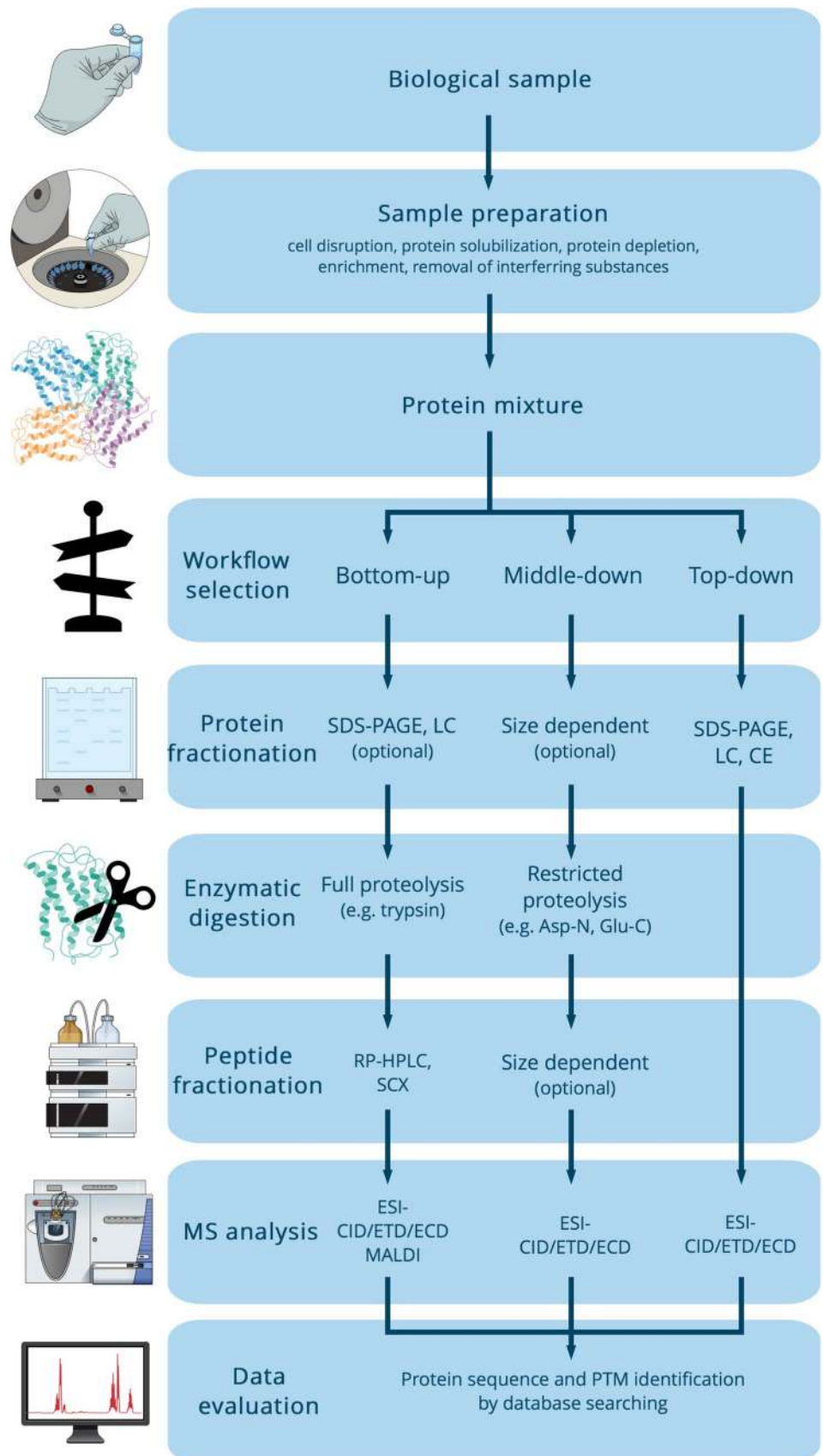
Mass Spectrometric Analysis of the Most Frequent Post-translational Modifications

Phosphorylation

Phosphorylation is widespread and often studied protein modification. Serine, threonine, and tyrosine residues are covalently modified by the enzymatic addition of a phosphate group. Phosphorylation is a reversible process, typically used to switch on and off various biological processes. The presence of the charged, hydrophilic group changes the structure of proteins, regulating multiple biochemical processes, for instance, cell cycle, metabolism and regulation of receptors [41, 42].

Phosphorylation of a given site typically occurs only in a (small) fraction of molecules. Phosphorylation analysis, therefore, requires both structure assignment (identification of the phosphorylation site) and determining the proportion of phosphorylated and non-phosphorylated analogs. Phosphoproteins constitute only a small fraction of the proteome, consequently enrichment strategies are essential for successful analysis. Several methods have been developed for this purpose, such as immunoaffinity chromatography, immobilized metal ion affinity chromatography (IMAC), and metal oxide affinity chromatography (MOAC) such as titanium dioxide (TiO₂) and zirconium dioxide (ZrO₂) chromatography [43–46]. IMAC and MOAC are based on the affinity of the phosphate groups for transition metal ions and metal oxides. However, these materials also show affinity for carboxyl groups, resulting in the isolation of acidic residues such as glutamate and aspartate-containing peptides alongside phosphorylated ones [47, 48]. Numerous strategies have been developed to overcome these specificity

Fig. 2 Typical proteomic workflows [33]



issues. Esterification of carboxyl groups prior to enrichment reduces the retention of acidic peptides. Organic acids having higher affinity for the stationary phase than carboxyl groups but lower than phosphate groups are employed as non-phosphopeptide excluders. Phosphopeptides can also be specifically eluted from the IMAC material using beta elimination, which removes the bond between the peptide and the phosphate group [48]. Although MOAC is shown to be more resistant to interfering compounds (e.g. salts, detergents) and more selective than IMAC in many cases, multiply phosphorylated residues may bind with high affinity to the metal oxide, which makes them dramatically difficult to elute. IMAC, therefore, provides improved coverage of multiphosphorylated peptides, and it is also able to enrich full-length phosphoproteins without any prior proteolytic digestion [48–50]. Immunoaffinity chromatography is predominantly employed in the enrichment of proteins phosphorylated on the tyrosine residue [51]. Considering that the data obtained with different enrichment methods only partly overlap each other (around 30%), combination of multiple methods is recommended to improve phosphoproteome coverage [52]. The enriched samples can be subsequently fractionated prior to MS analysis by multiple methods such as hydrophilic interaction chromatography (HILIC), RP-HPLC, and SCX [53–55].

There are two important issues that make phosphorylated peptide characterization by MS nontrivial. First, the sensitivity of a phosphopeptide is significantly lower than that of its non-phosphorylated analog [56]. Second, phosphopeptides typically fragment in CID by losing phosphoric acid, and therefore, identification of the phosphorylation site may be compromised [43]. Using ECD, ETD and/or special MS/MS techniques (such as MS³) may alleviate this problem and allow both sequence analysis and locating the phosphorylation site [57]. Phosphopeptide analysis may be performed using derivatization as well. Phosphoserine and phosphothreonine residues are selectively transformed into lysine analog; and subsequent proteolytic cleavage (using, e.g. trypsin or Lys-C) results in novel peptides [58]. Identifying these peptides using MS/MS and comparing it to the MS/MS of non-derivatized analog can be used to identify the phosphorylation site [46, 58].

Acetylation

Over the last few years, study of protein acetylation has become a major issue in PTM analysis. Proteins can be acetylated at the N terminus and at the ϵ -amino group of lysine residues, the latter having the greater biological relevance. Acetylation has been most extensively characterized for the regulation of histones, concluding that it plays an important role in cell cycle processes, such as gene expression and DNA repair [59, 60].

Since acetylation usually occurs at a very low stoichiometry, enrichment of the modified proteins and peptides is of critical relevance. Acetyllysine-specific antibodies are highly valuable for this purpose, and used as the initial purification step of very complex samples [61–63]. However, the method works properly only for peptides, as the accessibility of acetylated amino-acid residues is limited in the case of intact proteins. Moreover, there is no available antibody for the enrichment of N-terminal acetylation [64]. A number of methods are accessible for the prefractionation of histones including size exclusion chromatography (SEC), SDS-PAGE, and CE [65–67]. Acetylation increases the hydrophobic character of proteins and removes the positive charge from the N-terminal amino group. Acetylated residues, therefore, can be separated from their non-acetylated counterparts based on hydrophobicity (RP-HPLC), electrostatic interactions (SCX), or both (zwitterionic hydrophilic interaction liquid chromatography, ZIC-HILIC) [64, 68]. HILIC also has the added value of being able to separate methylated and acetylated histones, significantly decreasing sample complexity [69]. Another possibility for reducing sample complexity is derivatization by propionic anhydride, which converts lysine residues and the N-terminal amino group into propionyl amides providing a + 56 Da mass shift and protection from enzymatic digestion [37, 70].

Lysine acetylation is considered to be a stable PTM under mass spectrometry analysis, i.e. the functional group is retained on the protein using high-energy CID [23]. Acetylation can be identified through a 42.01 Da mass shift, in contrast with the unmodified variant. Although another modification called trimethylated lysine has a very similar weight (42.04 Da) to lysine acetylation, high-resolution mass spectrometers are capable to differentiate such modifications [71]. Since acetylation neutralizes the positive charge of lysine residue and blocks the tryptic cleavage, the absence of this characteristic cleavage can be another sign of acetylation [72].

O/N Glycosylation

Approximately, more than the half of human proteins bears simple or complex glycan side-chains, influencing protein folding and stability [73, 74]. There are two major types of glycan side-chains: (1) O-linked glycans attached to the hydroxy group of serine, threonine, tyrosine, while, (2) N-linked glycans are attached to the nitrogen of asparagine side-chains. In contrast to the synthesis of nucleic acids and proteins, glycosylation processes are not template directed, providing great heterogeneity for glycoconjugates. Consequently, a glycoprotein is a mixture of protein isoforms (glycoforms), instead of being a well-defined, exact chemical entity [75].

Complete characterization of glycoproteins and glycopeptides requires multiple analytical methods. In one approach, glycan side-chains are released from the protein by the application of an enzyme (e.g. peptide *N*-glycosidase F, PNGase F) or by reductive elimination. The former can be used in the case of *N*-glycans, and the latter for *O*-glycans [76–79]. When the sugars are released, they are typically separated from the protein mixture and derivatized. Replacing all reactive hydrogens with methyl groups (permethylation) is the most widespread method for this purpose, which enables detailed structural determination of complex glycans (e.g. branching sites, interglycosidic linkages, configurational isomers) [82]. As the procedure stabilizes sialic acid residues, permethylated derivatives show more effective ionization and more predictable fragmentation properties compared to their native counterparts [80–84]. Permethylated glycans are most often analyzed by MALDI-MS. Anthranilic acid (2-aminobenzoic acid, 2-AA) labeling of reducing carbohydrates is carried out in mild conditions, therefore, undesirable desialylation can be avoided [85]. 2-AA tag improves both chromatographic and mass spectrometric (in negative ion mode) detectabilities of glycans by acting as a chromophore, fluorophore and adding a negative charge to all the molecules. It also provides more informative fragments facilitating sequential analysis of saccharides [81, 85–88]. The derivatives can be subsequently separated by a number of methods, such as HILIC and CE prior to mass spectrometric analysis. CE is capable of separating positional isomers as well as differently linked glycans with high resolution and short analysis time, while HILIC tends to yield better retention time repeatability [89, 90]. It is also demonstrated that the techniques show notable complementarity in the glycoform profiling of therapeutic proteins [91].

An alternative to release glycan analysis is digestion of the protein (mixture) with proteolytic enzymes (e.g. trypsin). This yields a mixture of peptides and glycopeptides, and can be analyzed by CID- or ECD-based tandem mass spectrometry. Enzymatic digestion is often followed by enrichment techniques prior to the MS analysis. Glycopeptides may be enriched by boronate affinity chromatography, lectin affinity chromatography, HILIC, and electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) [75, 92–96]. Boronate affinity chromatography is able to capture glycan moieties through formation of borate diesters with vicinal diols of the sugar residues. This ensures selective enrichment of glycopeptides, even in complex mixtures. In contrast to boronate affinity chromatography, lectin affinity chromatography allows the enrichment of unique glycoproteins/peptides having specific glycan structure, which may be exceptionally useful in the exploration of disease-related, abnormal glycosylation pattern [97]. As glycosylation increases hydrophilicity of proteins and peptides, glycopeptides can be easily separated from non-glycosylated ones

by HILIC. ERLIC, in which positively charged functional groups are attached to the stationary phase, can be considered as a combination of hydrophilic interaction and ion-exchange chromatography. Sialylated moieties with negative charge are, therefore, highly retained, while positively charged peptides are easily eluted [98]. It is also reported that certain gradient HILIC methods can be converted into isocratic separations by ERLIC [99].

There are various mass spectrometric approaches for glycopeptide analysis, identifying the glycan structure, the glycosylation site and also the abundance ratio of various glycoforms. CID fragmentation often cleaves off the sugar residue, while ECD and ETD can yield information on the peptide sequence [75]. Ion mobility–mass spectrometry (IM-MS) is also a promising analytical tool in this field with the added value of being able to separate analytes based on their size. This technique has been successfully applied for the separation and characterization of isobaric polysaccharides and glycopeptides [100, 101].

Amidation

Approximately, half of the endocrine and neural peptides bear C-terminal amidation [102]. The presence of such modification is indispensable for signal transduction and receptor recognition [102, 103]. Furthermore, derivatization of amidated peptides with glycosylphosphatidylinositol is responsible for anchoring peptides and proteins to the cell membrane [104].

As in most cases in the field of PTMs, proteins that are presumably amidated at the C terminus can be separated or enriched by specifically engineered approaches based on the affinity of such proteins. For a long time, only radioimmunoassay (RIA) and immunoprecipitation (IP) have been applied for identification of amidated proteins [105, 106]. Separation of amidated peptides from their precursors can be carried out with RP-HPLC [105]. Weak cation-exchange chromatography (WCX) has been reported to be able to separate the α -amidated heavy chain of immunoglobulin G1 from other isoforms [107].

Amidation of a carboxyl group manifests in a 1 Da mass shift, which is very difficult to identify. The situation can be further complicated when the C-terminal amino acid is glutamine or asparagine, which cannot be distinguished from the amidated form of glutamate or aspartate, except when an auxiliary chemical or enzymatic procedure is performed. An approach has been developed for detecting C-terminal amidation by a combination of chemical derivatization and MALDI-TOF/TOF MS [86]. First step is the derivatization of the free carboxyl group at the C terminal, resulting in a methylamide structure [108, 109]. This results in a 13 Da mass increment observable in the spectrum, enabling the distinction between the amidated and the unmodified peptides.

Table 2 Separation/enrichment techniques, fragmentation methods, and commonly used derivatization methods of modified peptides and proteins [23, 43, 48, 58, 64, 70, 75, 82, 85, 97, 105, 107, 108, 114]

PTM type	Separation/enrichment techniques	Fragmentation methods	Commonly used derivatization methods
Phosphorylation	IMAC, MOAC, HILIC	CID, ECD, ETD	Transformation to lysine analogs
Acetylation	RP-HPLC, SDS-PAGE, CE, HILIC	CID	Propionyl amide formation at C terminal
O/N-linked glycosylation	Boronate affinity chromatography, lectin affinity chromatography, HILIC, ERLIC	CID, ECD, ETD	(Per-)methylation 2-AA labeling ^a
Amidation	WCX, RP-HPLC	CID	Amide formation at C terminal
Hydroxylation	RP-HPLC	CID, ECD, ETD	–

^aDerivatization of O/N-linked glycosylation is carried out at the released glycan level

This method was shown to be able to distinguish isobaric residues at the C terminal, such as Gln-OH and Glu-NH₂ or Asn-OH and Asp-NH₂, suggesting a wider application to investigate modified and unmodified C terminals of proteins [104].

Hydroxylation

Although hydroxylation can take place on several amino acids including lysine, histidine, asparagine, aspartate, and aromatic residues, the most commonly hydroxylated amino acid is proline, which makes up almost one-third of collagen protein [110–112]. Hydroxylation generally takes place at the γ -C atom, yielding 4-hydroxyproline (4-Hyp), which has the ability to stabilize the secondary structure of the protein by the powerful electronegative effect of the hydroxy group [112]. Hyp is also associated with the decreased availability of oxygen in the cellular environment, as the alpha subunits of hypoxia-induced factor (HIF) contain hydroxylated proline residues [113].

Mass spectrometric investigation of collagenous proteins is intricate due to the large number of proline residues and the high abundance of hydroxylation. These structural features result in a vast number of isobaric but differently modified peptides in a typical proteomic workflow, which usually co-elute during chromatographic separation and provide chimeric spectra troublesome to elucidate. Moreover, the nominal mass of Hyp is 113 Da, which is the same as leucine and isoleucine. To overcome such difficulties, a mass spectrometer with high mass accuracy and abundant product ions from the tandem MS experiment are required [114, 115]. Unfortunately, MSⁿ sequence analysis is ruined by the abnormal fragmentation of Pro- and Hyp-rich proteins and peptides. These residues promote characteristic and dominant cleavages instead of those nonselective ones, which may provide sequence information. This phenomenon is the so-called “proline-effect”, by which the identification of collagens and other proline-rich proteins are greatly hampered [114, 116, 117]. An approach has been suggested that is based

on the application of five different proteolytic enzymes to increase sequence coverage of collagenous proteins. Thereafter, the analytes were separated by nano-LC and introduced into a linear ion trap–orbitrap instrument. Differently modified peptides with the same sequence were eluted and ionized together producing chimeric mass spectra. As ETD-induced fragmentation did not yield sufficient amount of data, additional dissociation techniques (CID and HCD) were also applied. In general, multistage activation made the identification of PTM-carrying residues possible, however, in some cases, the exact localization of the modification has remained unknown [114].

Table 2 summarizes the separation/enrichment techniques, fragmentation methods, and commonly used derivatization methods of modified peptides and proteins.

Conclusion and Outlook

Formation of post-translational modifications is the key generator of protein diversity. To understand the physiological/pathophysiological processes in our organism and be able to produce better quality and more efficient biopharmaceutical products and diagnostic procedures, further knowledge needs to be gained in the field of proteomics and PTM mapping. Mass spectrometry-based proteomic approaches has emerged as a powerful device for screening and characterizing PTMs. Although the current technology is unable to provide a complete picture of the modified proteome, future proteomics will rely heavily on continued improvements of mass spectrometry-based proteomic strategies.

Acknowledgment Open access funding provided by Semmelweis University (SE).

Funding This work was financed by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the molecular biology thematic programme of Semmelweis University. The research was also supported

by the National Research Development and Innovation Office (NKFIH K-119459).

Compliance with Ethical Standards

Conflict of Interest The authors declare that there are no conflicts of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Sridhar GR (2001) *J Assoc Physicians India* 49:995–998
- Muñoz J, Heck AJ (2014) *Angew Chem Int Ed Engl* 41:10864–10866
- Jensen ON (2004) *Curr Opin Chem Biol* 8:33–41
- Roberts GC, Smith CWJ (2002) *Curr Opin in Chem Biol* 6:375–383
- Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J (2000) *Molecular cell biology*. Freeman & Co., New York
- Duan G, Walther D (2015) *PLoS Comput Biol* 11:e1004049
- Khoury GA, Baliban RC, Floudas CA (2011) *Sci Rep* 1:90
- Wesche J, Kühn S, Kessler BM, Salton M, Wolf A (2017) *Cell Mol Life Sci* 74:3305–3315
- Miranda M, Sorkin A (2007) *Mol Interv* 7:157–167
- Liu YD, Goetze AM, Bass RB, Flynn GC (2011) *J Biol Chem* 286:11211–11217
- Huang BY, Chen PC, Chen BH, Wang CC, Liu HF, Chen YZ, Chen CS, Yang YS (2017) *Anal Chem* 89:3278–3284
- Hounsell EF (1994) *Adv Carbohydr Chem Biochem* 50:311–350
- Barret AW, Speight PM (1996) *J Clin Pathol* 49:565–569
- Lacunza I, Sanz J, Diez-Masa JC, de Frutos M (2006) *Electrophoresis* 27:4205–4214
- Hounsell EF, Young M, Davies MJ (1997) *Clin Sci* 93:287–293
- Mills P, Mills K, Clayton P, Johnson A, Whitehouse D, Winchester B (2001) *Biochem J* 359:249–257
- Harvey DJ (2001) *Proteomics* 1:311–328
- Fournier T, Medjoubi-N N, Porquet D (2000) *Biochim Biophys Acta* 1482:157–171
- Cecilian F, Pocacqua V (2007) *Curr Protein Pept Sci* 8:91–108
- Schraen-Maschke S, Sergeant N, Dhaenens CM, Bombois S, Deramecourt V, Caillet-Boudin ML, Pasquier F, Muraige CA, Sablonnière B, Vanmechelen E, Buée L (2008) *Biomark Med* 2:363–384
- Arito M, Nagai K, Ooka S, Sato T, Takakuwa Y, Kurokawa MS, Sase T, Okamoto K, Suematsu N, Kato T (2015) *Clin Exp Rheumatol* 33:877–886
- Amann T, Schmieder V, Fastrup Kildegaard H, Borth N, Andersen MR (2019) *Biotechnol Bioeng* 1–19
- Mann M, Jensen ON (2003) *Nat Biotechnol* 21:255–261
- Boyd RK (1994) *Mass Spectrom Rev* 13:359–410
- Signor L, Boeri Erba E (2013) *J Vis Exp* 79:e50635
- Cramer R, Pirkel A, Hillenkamp F, Dreisewerd K (2013) *Angew Chem Int Ed Engl* 52:2364–2367
- Ho CS, Lam CWK, Chan MHM, Cheung RCK, Law LK, Lit LCW, Ng KF (2003) *Suen MWM Tai HL. Clin Biochem Rev* 24:3–12
- Miladi M, Harper B, Solouki T (2013) *J Am Soc Mass Spectrom* 24:1755–1766
- Mikesh LM, Ueberheide B, Chi A, Coon JJ, Syka JEP, Shabanowitz J, Hunt DF (2006) *Biochim Biophys Acta* 1764:1811–1822
- Feist P, Hummon AB (2015) *Int J Mol Sci* 16:3537–3563
- Amunugama R, Jones R, Ford M, Allen D (2013) *Adv Wound Care* 2:549–557
- Yates JR, Ruse CI, Nakorchevsky A (2009) *Annu Rev Biomed Eng* 11:49–79
- Scherp P, Ku G, Coleman L, Kheterpal I (2011) *Methods Mol Biol* 702:163–190
- Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR (2013) *Chem Rev* 113:2343–2394
- Wolters DA, Washburn MP, Yates JR (2001) *Anal Chem* 73:5683–5690
- Silva AMN, Vitorino R, Domingues MRM, Spickett CM, Domingues P (2013) *Free Radic Biol Med* 65:925–941
- Moradian A, Kalli A, Sweredoski MJ, Hess S (2014) *Proteomics* 14:489–497
- Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SMM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL (2011) *Nature* 480:254–258
- Lorenzatto KR, Kim K, Ntai I, Paludo GP, Camargo de Lima J, Thomas PM, Kelleher NL, Ferreira HB (2015) *J Proteome Res* 14:4805–4814
- Durbin KR, Fornelli L, Fellers RT, Doubleday PF, Narita M, Kelleher NL (2016) *J Proteome Res* 15:976–982
- Karasev DA, Veselova DA, Veselovsky AV, Sobolev BN, Zgoda VG, Archakov AI (2018) *Proteins* 86:13–20
- Humphrey SJ, James DE, Mann M (2015) *Trends Endocrinol Metab* 26:676–687
- Thingholm TE, Larsen MR (2016) *Methods Mol Biol* 1355:147–160
- Bergström Lind S, Hagner-McWhirter S, Elfineh L, Molin M, Jorsback A, Ohman J, Pettersson U (2010) *Biochem Biophys Res Commun* 401:581–585
- Fila J, Honys D (2017) *Methods Mol Biol* 1669:265–274
- Palumbo AM, Smith SA, Kalcic CL, Dantus M, Stemmer PM, Reid GE (2011) *Mass Spectrom Rev* 30:600–625
- Ke M, Shen H, Wang L, Luo S, Lin L, Yang J, Tian R (2016) *Adv Exp Med Biol* 919:345–382
- Beltran L, Cutillas PR (2012) *Amino Acids* 43:1009–1024
- Bolanos-Garcia VM, Davies OR (2006) *Biochim Biophys Acta* 1760:1304–1313
- Shetty J, Sinville R, Shumilin IA, Minor W, Zhang J, Hawkinson JE, Georg GI, Flickinger CJ, Herr JC (2016) *Protein Expr Purif* 121:88–96
- Rush J, Moritz A, Lee KA, Guo A, Goss VL, Spek EJ, Zhang H, Zha XM, Polakiewicz RD, Comb MJ (2005) *Nat Biotechnol* 23:94–101
- Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R (2007) *Nat Methods* 4:231–237
- Engholm-Keller K, Larsen MR (2016) *Methods Mol Biol* 1355:161–177
- Batth TS, Olsen JV (2016) *Methods Mol Biol* 1355:179–192
- Alpert AJ, Hudecz O, Mechtler K (2015) *Anal Chem* 87:4704–4711
- Gafken PR, Lampe PD (2006) *Cell Commun Adhes* 13:249–262
- Potel CM, Lemeer S, Heck AJR (2019) *Anal Chem* 91:126–141

58. Knight ZA, Schilling B, Row RH, Kenski DM, Gibson BW, Shokat KM (2003) *Nat Biotechnol* 21:1047–1054
59. Averbek NB, Durante M (2011) *J Cell Physiol* 226:962–967
60. Verdin E, Ott M (2015) *Nat Rev Mol Cell Biol* 16:258–264
61. Kim SC, Sprung R, Chen Y, Xu Y, Ball H, Pei J, Cheng T, Kho Y, Xiao H, Xiao L, Grishin NV, White M, Yang XJ, Zhao Y (2006) *Mol Cell* 23:607–618
62. Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M (2009) *Science* 325:834–840
63. Choudhary C, Weinert BT, Nishida Y, Verdin E, Mann M (2014) *Nat Rev Mol Cell Biol* 15:536–550
64. Diallo I, Seve M, Cunin V, Minassian F, Poisson JF, Michelland S, Bourgoin-Voillard S (2019) *Expert Rev Proteomics* 16:139–159
65. Klinker H, Haas C, Harrer N, Becker PB, Mueller-Planitz F (2014) *PLoS One* 9:e104029
66. Cong X, Held JM, DeGiacomo F, Bonner A, Chen JM, Schilling B, Czerwiec GA, Gibson BW, Ellerby LM (2011) *Mol Cell Proteomics* 10:M111.009829
67. Sarg B, Faserl K, Kremser L, Halfinger B, Sebastiano R, Lindner HH (2013) *Mol Cell Proteomics* 12:2640–2656
68. Papazyan R, Taverna SD (2013) *Methods Mol Biol* 981:103–113
69. Jiang T, Hoover ME, Holt MV, Freitas MA, Marshall AG, Young NL (2018) *Proteomics* 18:e1700442
70. Lin S, Wein S, Gonzales-Cope M, Otte GL, Yuan ZF, Afjehi-Sadat L, Maile T, Berger SL, Rush J, Lill JR, Arnott D, Garcia BA (2014) *Mol Cell Proteomics* 13:2450–2466
71. Zee BM, Garcia BA (2013) *Methods Mol Biol* 981:1–11
72. Weinert BT, Satpathy S, Hansen BK, Lyon D, Jensen LJ, Choudhary C (2017) *Mol Cell Proteomics* 16:759–769
73. Li F, Li C, Wang M, Webb GI, Zhang Y, Whisstock JC, Song J (2015) *Bioinformatics* 31:1411–1419
74. Joao HC, Dwek RA (1993) *Eur J Biochem* 218:239–244
75. Schiel JE (2012) *Anal Bioanal Chem* 404:1141–1149
76. Holst S, van Pelt GW, Mesker WE, Tollenaar RA, Belo AI, van Die I, Rombouts Y, Wührer M (2017) *Methods Mol Biol* 1503:185–196
77. Jensen PF, Comamala G, Trelle MB, Madsen JB, Jørgensen TJ, Rand KD (2016) *Anal Chem* 88:12479–12488
78. Hanisch FG, Müller S (2009) *Methods Mol Biol* 534:107–115
79. Morelle W, Faid V, Chirat F, Michalski JC (2009) *Methods Mol Biol* 534:5–21
80. Zhou S, Wooding K, Mechref Y (2018) *Methods Mol Biol* 1503:83–96
81. Ruhaak LR, Zauner G, Huhn C, Bruggink C, Deelder AM, Wührer M (2010) *Anal Bioanal Chem* 397:3457–3481
82. Kang P, Mechref Y, Klouckova I, Novotny MV (2005) *Rapid Commun Mass Spectrom* 19:3421–3428
83. Lin Z, Lubman DM (2013) *Methods Mol Biol* 1007:289–300
84. Zhou S, Veillon L, Dong X, Huang Y, Mechref Y (2017) *Analyst* 142:4446–4455
85. Anumula KR (2014) *Anal Biochem* 457:31–37
86. Neville DCA, Alonzi DS, Butters TD (2012) *J Chrom A* 1233:66–70
87. Kamoda S, Ishikawa R, Kakehi K (2006) *J Chrom A* 1133:332–339
88. Imre T, Kremmer T, Héberger K, Molnár-Szöllosi E, Ludányi K, Pócsfalvi G, Malorni A, Drahos L, Vékey K (2008) *J Proteomics* 71:186–197
89. Lu G, Crihfield CL, Gattu S, Veltri LM, Holland LA (2018) *Chem Rev* 118:7867–7885
90. Kok MG, Somsen GW, de Jong GJ (2015) *Talanta* 132:1–7
91. Domínguez-Vega E, Tengattini S, Peintner C, van Angeren J, Temporini C, Haselberg R, Massolini G, Somsen GW (2018) *Talanta* 184:375–381
92. Chen J, Li X, Feng M, Luo K, Yang J, Zhang B (2017) *Anal Bioanal Chem* 409:519–528
93. Zhu F, Trinidad JC, Clemmer DE (2015) *J Am Soc Mass Spectrom* 26:1092–1102
94. Zacharias LG, Hartmann AK, Song E, Zhao J, Zhu R, Mirzaei P, Mechref Y (2016) *J Proteome Res* 15:3624–3634
95. Cao L, Yu L, Guo Z, Li X, Xue X, Liang X (2013) *J Chrom A* 1299:18–24
96. Adav SS, Hwa HH, de Kleijn D, Sze SK (2015) *J Prot Res* 14:2828–2838
97. Ahn YH, Kim JY, Yoo JS (2015) *Mass Spectrom Rev* 34:148–165
98. Zhu R, Zacharias L, Wooding KM, Peng W, Mechref Y (2017) *Methods Enzymol* 585:397–429
99. Alpert AJ (2008) *Anal Chem* 80:62–76
100. Gray CJ, Thomas B, Upton R, Migas LG, Evers CE, Barran PE, Flitsch SL (2016) *Biochim Biophys Acta* 1860:1688–1709
101. Morrison KA, Clowers BH (2018) *Curr Opin Chem Biol* 42:119–129
102. Czyzyk TA, Ning Y, Hsu MS, Peng B, Mains RE, Eipper BA, Pintar JE (2005) *Dev Biol* 287:301–313
103. Chufán EE, De M, Eipper BA, Mains RE, Amzel LM (2009) *Structure* 17:965–973
104. Kuyama H, Nakajima C, Nakazawa T, Nishimura O, Tsunasawa S (2009) *Proteomics* 9:4063–4070
105. Mueller GP, Driscoll WJ (2008) *Methods Mol Biol* 446:67–84
106. Yamaguchi H, Sasaki K, Satomi Y, Shimbara T, Kageyama H, Mondal MS, Toshinai K, Date Y, González LJ, Shioda S, Takao T, Nakazato M, Minamino N (2007) *J Biol Chem* 282:26354–26360
107. Johnson KA, Paisley-Flango K, Tangarone BS, Porter TJ, Rouse JC (2007) *Anal Biochem* 360:75–83
108. Nakazawa T, Yamaguchi M, Nishida K, Kuyama H, Obama T, Ando E, Okamura TA, Ueyama N, Tanaka K, Norioka S (2004) *Rapid Commun Mass Spectrom* 18:799–807
109. Yamaguchi M, Oka M, Nishida K, Ishida M, Hamazaki A, Kuyama H, Ando E, Okamura TA, Ueyama N, Norioka S, Nishimura O, Tsunasawa S, Nakazawa T (2006) *Anal Chem* 78:7861–7869
110. Zurlo G, Guo J, Takada M, Wei W, Zhang Q (2016) *Biochim Biophys Acta* 1866:208–220
111. Zheng S, Bai X, Tian S, Wang G, Zhai G, Guo Z, Bi W, Shen L, Zhang K (2016) *Rapid Commun Mass Spectrom* 30:185–189
112. Holmgren SK, Bretscher LE, Taylor KM, Raines RT (1999) *Chem Biol* 6:63–70
113. D'Angelo G, Duplan E, Boyer N, Vigne P, Frelin C (2003) *J Biol Chem* 278:38183–38187
114. Yang C, Park AC, Davis NA, Russell JD, Kim B, Brand DD, Lawrence MJ, Ge Y, Westphall MS, Coon JJ, Greenspan DS (2012) *J Biol Chem* 287:40598–40610
115. Chicooree N, Unwin RD, Griffiths JR (2015) *Mass Spectrom Rev* 34:595–626
116. Raulfs MD, Brecci L, Bernier M, Hamdy OM, Janiga A, Wysocki V, Poutsma JC (2014) *J Am Soc Mass Spectrom* 25:1705–1715
117. Grewal RN, El Aribi H, Harrison AG, Siu KWM, Hopkinson AC (2004) *J Phys Chem B* 108:4899–4908

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.