Post-translational Modifications and Mass Spectrometry Detection

André Silva\textsuperscript{a}, Rui Vitorino\textsuperscript{a}, M. Rosário M. Domingues\textsuperscript{a}, Corinne M. Spickett\textsuperscript{b}, Pedro Domingues\textsuperscript{a}\textsuperscript{*}

\textsuperscript{a} Mass Spectrometry Centre, QOPNA, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal

\textsuperscript{b} School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, United Kingdom

\textsuperscript{*} Author to whom correspondence should be addressed:

Pedro Domingues

e-mail: p.domingues@ua.pt

Phone:+351 234 370 699

Fax:+ 351 234 370 084

Keywords: Post-translational modifications; Mass spectrometry; Literature review

Abstract

In this review, we provide a comprehensive bibliographic overview of the role of mass spectrometry and the recent technical developments in the detection of post-translational modifications (PTMs). We briefly describe the principles of mass spectrometry for detecting PTMs and the protein and peptide enrichment strategies for PTM analysis, including phosphorylation, acetylation and oxidation. This review presents a bibliographic overview of the scientific achievements and the recent technical development in the detection of PTMs is provided. In order to ascertain the state of the art in mass spectrometry and proteomics methodologies for the study of PTMs, we analyzed all the PTM data introduced in the Universal Protein Resource (UniProt) and the literature published in the last three years. The evolution of curated data in UniProt for proteins annotated as being post-translationally modified is also analyzed. Additionally, we have undertaken a careful analysis of the research articles published in the years 2010 to 2012 reporting the detection of PTMs in biological samples by mass spectrometry.

The use of proteomics for detecting and quantifying post-translational modifications (PTM)

The development of two-dimensional gel electrophoresis in 1975 by O'Farrell (1), together with the seminal work by Klose (2) and Scheele (3), resulted in the publication of what are now considered early applications of proteomics (4-6) (reviewed by (7)). The concept of the proteome, defined as protein-based gene expression analysis, was introduced by Wilkins et al (8). Later, James introduced
the term proteomics, as an analogy with genomics (9). The actual definition has evolved during the
time, but it can be understood that “Proteomics includes not only the identification and
quantification of proteins, but also the determination of their localization, modifications,
interactions, activities, and, ultimately, their function” (10). Although proteomics involves the use of
various technologies, including nuclear magnetic resonance (NMR), light and electron microscopy,
Fourier transform infrared spectroscopy, array and chip experiments, and others, mass spectrometry
(MS) has become an essential technology.

In fact, in recent years mass spectrometry has evolved dramatically, and is now considered a key
technology for identifying and quantifying proteins and post-translational modifications (PTMs) and
for detecting protein–protein interactions. This new role of mass spectrometry in the biological field
is the result of the development of new ionization techniques; faster, more sensitive and accurate
analyzers, development of the fragmentation methods for tandem mass spectrometry, technology
advances in separation technologies and rapid methods for mass spectrometry protein analysis. The
following section briefly describes a user’s perspective of the theoretical background of mass
spectrometry and the mass spectrometry techniques used for identifying and quantifying PTMs.

a) Principles of mass spectrometry for detecting post-translational modifications

Mass spectrometry

Mass spectrometry (MS) is an analytical technique that involves the study in the gas phase of
ionized molecules, aiming to determine the molecular weight of compounds, to characterize their
structure and to determine their abundance. Several reviews have been published recently,
describing in detail this technique and its application in proteomics (11-15). Herein, the general
concepts necessary for understanding the use of mass spectrometry for detecting PTMs will be
briefly reviewed. Briefly, the MS workflows generally consist of ionization of a sample in an ion
source, separation of the ionized molecules according to their mass-to-charge ratio in an analyzer,
detection of the ionized molecules in a detector, and generation of a mass spectrum. Tandem mass
spectrometers have the additional capability of selecting an ion and fragmenting it in order to obtain
structural information (MS/MS).

Several ionization methods have been developed for analyzing organic compounds (13, 16).
However, the development of two soft ionization techniques capable of ionizing non-volatile and
thermo-unstable biological compounds, namely electrospray ionization (ESI)(17) and matrix-assisted
laser desorption/ionization (MALDI) (18, 19), allowed protein analysis using MS. For a detailed view
of the ionization mechanisms, the interested reader is referred to specialist reviews for ESI (20) and
MALDI (21). Although both methods allow peptides to be ionized with high sensitivity, there are
significant differences between them: first of all, ESI seems to favor the observation of hydrophobic
peptides (22, 23) whereas MALDI seems to favor the observation of basic and aromatic peptides (22,
24, 25); additionally, MALDI seems to discriminate positively Arg-containing peptides, while ESI is less
affected by the presence of this amino acid residue(26, 27). This can be an important factor when
using proteolytic enzymes other than trypsin (28). Moreover, due to the occurrence of interference
ions generated by the matrix signal, MALDI is not well suited to the observation of low molecular
weight ions (<800 Da). Nevertheless, MALDI is usually considered to be less adversely affected by salt
contaminants. For this reason, this has been considered the method of choice for analyzing peptides
recovered from electrophoresis gels by peptide map fingerprinting (PMF) (29). One important factor that can affect the ionizing performance either of MALDI or ESI sources is ion suppression. This effect has been described in early soft ionization mechanistic studies (30-34) and it is due to sample matrix, co-eluting compounds, and interferences, also known as “cross-talks” (35, 36). Ion suppression is one of the main reasons why sample quality (purity) and matrix complexity are very important aspects in mass spectrometry analysis. Although ion suppression has deleterious effects on both ESI and MALDI, the relative insensitivity of the later ionization method to this effect might be an advantage (37). Nevertheless, aiming at higher sensitivity, nowadays most proteomics experimental approaches include a sample pre-fractionation step for peptide separation.

One of the most important aspects of PTM discovery is the consideration of protein sequence coverage. In fact, the very high dynamic range of protein concentrations in a human cell (at least seven orders of magnitude) (38), poses very important difficulties for the identification and quantification of proteins in general and low abundance proteins in particular. This high range highlights the need for the fractionation and enrichment of proteins and the importance of the development of rapid and sensitive methods (see below). As we will discuss later, most proteomics researchers perform sample fractionation and preconcentration using nano LC chromatography. This approach is used to increase protein identification rates and sequence coverage. A complete discussion on the variability of protein sequence coverage rate using mass spectrometry must also include a consideration of analyzers and MS/MS fragmentation methods. In addition, the different discrimination of MALDI and ESI peptides also contributes to different protein coverage rates. Several authors have reported different results when using these two ionization methods (22, 37, 39-41). These results globally show that ESI allows a slightly better protein coverage rate. Also, these studies suggest that, in order to enhance the protein coverage rate, both ionization methods should be employed. A comprehensive overview and discussion on protein coverage is given in reference (42).

When selecting a mass spectrometer for proteomics and PTM detection, several parameters should be considered. These include speed of acquisition, resolving power, mass accuracy, sensitivity, dynamic range and methods of fragmentation. The choice of a particular instrument is then dependent on the analysis of all these instrumental factors, plus the acquisition and maintenance costs and the available software package(s). Nevertheless, it is now considered that high resolving power and mass accuracy are essential for proteomics and PTM analysis (43, 44). This limits the instruments to three principal tandem mass spectrometry platforms: quadrupole time-of-flight (QqTOF), time-of-flight/time-of-flight (TOT/TOF) and ion trap/orbitrap (IT/orbitrap) mass spectrometers. We have excluded from this analysis an obvious alternative, FT-ICR instruments, because of their relatively low scan rate, high maintenance requirements as well as acquisition price. All these platforms have advantages and disadvantages (45, 46) but, as will be described later, in recent times, the orbitrap (47) has acquired a prominent role in proteomics in general and in PTM analysis in particular. The orbitrap is usually used for acquisition of low scan rate high resolution and accuracy MS data, and is normally coupled with an ion trap for high speed low resolution and accuracy MS experiments (IT/orbitrap). Nonetheless, with recent developments in instrumentation towards higher resolution, mass accuracy and speed in all platforms but especially in ion traps and in QqTof instruments, this scenario is expected to change, as demonstrated by some recent studies (48, 49).
The default method of fragmentation in mass spectrometry has been collision induced dissociation (CID) (50, 51). CID is based on the collision of selected molecular ions with a neutral atom or molecule. Part of the collision energy is converted into vibrational energy and fragmentation of the molecular ion can occur. When analyzing peptides, these fragment ions provide information on their amino acid composition and sequence, since the protonated amide linkages are favored for cleavage (52). However, it is important to bear in mind that when peptides contain PTMs, the preferred dissociation pathways can be different. Several common and important PTMs are known to be labile, such as sulfation, glycosylation and phosphorylation (pSer, pThr), and thus neutral loss of the modification can be the predominant fragmentation pathway (53). As a consequence, tandem mass spectra of PTMs containing peptides might contain inadequate peptide fragmentation and diagnostic sequence ion information for the localization of the modification site. To this end, it is worth noting the role of data-dependent neutral loss MS\(^3\) tandem mass spectrometry for identifying phosphopeptides (54) and glycopeptides (55); diagnostic neutral losses for oxidized cysteine and methionine residues have also been reported, as reviewed by Spickett and Pitt (56). Electron capture dissociation (ECD) is an alternative method for peptide ion dissociation introduced by McLafferty and coworkers in the late 90s (57). In this method, fragmentation occurs upon the capture of a low-energy electron of multiply charged ions trapped in an FT-ICR-MS instrument. This method induces cleavage of peptide backbones primarily at the N-C\(_\alpha\) bond rather than at the amide linkage, and early on it was found to allow the analysis of sulfated (58), glycosylated (59) and phosphorylated (60) peptides. More recently, electron transfer dissociation (ETD) was reported by Hunt and co-workers (61). In this method, fragmentation is induced by electron transfer to multiply protonated peptides. Electron transfer dissociation is usually performed using ion-trap mass spectrometers but it can also be performed in QTOF instruments (62). As described for ECD, the neutral loss of labile PTMs is not the preferred fragmentation pathway (reviewed in (63)), particularly in phosphorylated and glycated peptides (64). Several studies have shown that there is little overlap in peptide identifications between ETD and CID. ETD is a more informative fragmentation method than CID, particularly for charge states greater than 2 (65-67). This strongly suggests the benefit of using both methods for PTM discovery.

High-energy collision dissociation (HCD) is an alternative beam type tandem collision induced fragmentation method that has been implemented for ion trap mass spectrometers (68). Originally, this method was implemented in Orbitrap platforms with a C-tap, but has also been used to induce fragmentation in the ion injection pathway of ion trap mass spectrometers with a standard atmospheric inlet (69). HCD has the advantage of eliminating the low-mass cutoff of fragment ions in ion trap instruments and the ability to produce and detect immonium ions. Also, it has been reported that it improves the identification of PTMs, mainly due to the fact that the coverage of peptide sequence is higher in HCD (70, 71). The use of these three complementary fragmentation methods (CID, ETD and HCD) has been shown to increase the number of peptide and PTM identifications and to provide greater protein coverage (70, 72, 73).

Altogether, the data herein discussed has shown the importance of using complementary instrumental approaches for better protein sequence coverage and for PTM’s identification. In fact, until now, no single instrumental approach has been shown to possess determinant advantages over the others.

Detection of PTMs
Detection of PTM’s by mass spectrometry relies on the tryptic peptide mass alteration, which results from the chemical modification occurring in the amino acid residue side chain. However, for the unequivocal assignment of a given modification site, tandem mass spectrometry (MS/MS) experiments are necessary. In these experiments, it is necessary that the mass shift detected in the precursor ions (peptide from tryptic digestion) is also observed in the fragment ions carrying the modified amino acid residue.

One essential aspect of PTM identification is the need for high quality samples, i.e. high purity with low degradation and low abundance of contaminants, through the use of competitive separation techniques for intact proteins or peptides. As stated before, the very high dynamic range of protein concentrations in a cell poses serious difficulties for the identification of PTMs. Also, it is difficult or impossible to devise a strategy allowing the separation and analysis of all the structurally diverse protein isoforms, because of the number of possible combinations of PTM locations within a protein with several PTMs. To understand the complexity of the problem, Thelen and Miernyka have shown a virtual comparison of a liquid chromatography profile of a peptide containing zero to three PTMs (46), highlighting the need to use a combination of strategies to improve the identification of PTMs. Nowadays, several approaches are used for identifying and quantifying proteins and PTMs, which have been conveniently reviewed in several high quality papers (45, 53, 74, 75).

There are several strategies of PTM identification using mass spectrometry. These are essentially the same as those used for protein identification and have been summarized in figure 1: bottom-up, top-down and middle-down approaches.
Figure 1 - Overview of proteomic analysis for PTM detection. M corresponds to a modified residue. The right-hand panel shows the instrumental approaches that can be interfaced with different stages of samples manipulation.

The bottom-up approach is the traditional proteomic approach: for protein identification and PTM discovery, fractionated proteins are proteolytically digested and peptides are analyzed by liquid chromatography (LC) mass spectrometry (MS) and tandem mass spectrometry (MS/MS) (figure 1) (76). As mentioned before, although the sample fractionation and preconcentration using nano LC chromatography is important, low-complexity samples can be analyzed rapidly without peptide separation, especially when using MALDI-MS. In this review, studies in which the protein (or small group of proteins) of interest is purified or isolated prior to the proteomics analysis have been designated as “focused” studies. These purification processes usually include 1D or 2D gel electrophoresis methods. When an unfractionated entire proteome of a cell or tissue is analyzed, this approach is referred to as shotgun proteomics (77) (recently reviewed by Yates (78)). However, in this case, sample fractionation and preconcentration using nano LC chromatography, frequently using 2D liquid chromatography, is essential. Several workflows have been developed, following the pioneering work of Yates and co-workers in an approach named multidimensional protein identification technology (MudPIT) (79). Recently, the shotgun approach was used in the analysis of yeast proteomes, allowing the identification of 2990 yeast proteins corresponding to 35,155 sequence unique peptides, and average sequence coverage rate of 18% (80). Although this workflow allows high throughput protein identification, the assignment of PTM sites is compromised due to the MS dynamic range and the typical low sequence coverages that are obtained. Peptide enrichment (see below) decreases this problem and allows the shotgun approach to be applied to the identification and quantification of PTMs. For example, using titanium dioxide enrichment, Mann and co-workers have been able to quantify 6027 proteins and 20,443 unique phosphorylation sites in HeLa S3 cells (81). Within the bottom-up approach, several strategies for targeted detection of PTMs have been or are being developed, involving the use of neutral loss or precursor ion scanning routines to detect fragmentations diagnostic for the presence of PTMs, although as yet most have not been fully tested in proteomics of biological samples. Nevertheless, significant progress has been made, for example in the detection of phosphorylation sites by a combination of neutral loss and targeted product ion scanning (reviewed in (82)). Also, strategies for detecting the oxidation PTMs have been developed: The oxidation of sulphur in cysteine has been detected using characteristic neutral losses scanning ((83-85). Also, MS³ precursor ion scanning has been used to detect chlorotyrosine, hydroxytyrosine and hydroxytryptophan residues (86). MS³ (87) and MS² (88) scanning for neutral losses were also applied to the detection of adducts between HNE and proteins. Other methods using precursor-scanning-like analysis on a hybrid instrument with high resolution in the second analyzer for detection of oxidative modifications to proteins are also in development (48).

The top-down strategy was initially developed in McLafferty’s laboratory (89-91) and is now considered one the most promising proteomic approaches in terms of determination of the relative occupancy of modification sites. In this approach, intact proteins are ionized by electrospray ionization and fragmented, usually using ECD or ETD, and the fragment ions analyzed using high resolution analyzers such as FT-ICR or orbitraps (reviewed in (14)). Although these are the preferred platforms for top-down analysis, other ionization methods: MALDI (92); fragmentation methods: CID
(93) or infrared multiphoton dissociation (IRMPD) (94)); and analyzers: QTOF (95) or IT (96)), have also been used. One important aspect in this approach is that the proteins under study are purified to a level suitable for analysis by mass spectrometry, usually in the range of 150 to 250 fmol (97). As proof of concept for top-down large scale use in protein and PTM discovery, Kelleher and co-workers have analyzed over 3,000 intact proteins from HeLa S3 cells and detected 645 phosphorylations, 538 lysine acetylations, 158 methylations, 19 lipid/terpenes and 5 hypusines (98).

The middle-down approach can be considered as a variant of the top-down approach. In this approach, proteins are subjected to soft proteolysis, usually using AspN and GluC, and the large peptides obtained (typically with masses between 3K Da and 10KDa) are sequenced by tandem mass spectrometry using platforms similar to those described for top-down. Curiously, one of the first applications of the top-down approach (99), according to this definition, would now be considered a middle-down experiment. This approach was first explored by Forbes et al. (100) and allows detection of PTMs with high efficiency, by combining the advantages of top-down and bottom-up approaches. One important demonstration of the efficacy of this approach was the work of Garcia and co-workers, where over 200 H3.2 and 70 H4 histone forms were reported (101).

**Data analysis for PTM detection**

High-throughput proteomics allows the simultaneous identification of thousands of modified peptides. Of all the work-flow tasks, data processing, analysis and presentation are probably the most challenging aspects of proteomics. Unless targeted approaches are used, the characterization of PTMs using MS/MS data involves exhaustive searches of all potential combinations of mass shift for each identified peptide from a protein database (102-104). Several tools are most commonly used to perform this task, although the most used are Mascot (Matrixscience) (105), Sequest (106), Maxquant (107) and Paragon (108).

Recently, different protein database search algorithms for identifying candidate PTM peptides from MS/MS spectra have been developed. An important evolution in PTM detection algorithms enables the interpretation of experimental data combined with referenced or predicted PTM sites. Several groups have been working on the computational prediction of PTM sites. For example, GPS (Group-based Prediction System) is a freely available tool for prediction of kinase target sites for phosphorylation (109). In silico methods have also been developed for identifying protein oxidation products in biological systems (110). Other more frequently used potential protein post-translational modification predictions tools include the prediction of phosphorylation (NetPhos, NetPhosK (111, 112) PredPhospho (113) and glycosylation (NetOGlyc (114), NetNGlyc, DictyOGlyc, YinOYang (115), GlycoPP (116)) and other modifications (reviewed in (117)). Other approaches use consensus-based approaches, combining several signature recognition methods to scan a given query protein sequence against observed protein signatures (104). PROSITE is a typical example of a bioinformatic tool that scans a query sequence to find consensus patterns for several types of PTMs (118, 119). However, it is important to note that the presence of a consensus sequence does not necessarily equate to the presence of the modification.

Although the development of new user-friendly bioinformatics tools for the large-scale analysis of PTMs has allowed significant progress in proteomics and PTM discovery, manual validation of the MS/MS data is still necessary. However, most large-scale proteomics PTMs reports, although providing information about the computational constraints applied to data interpretation, do not
manually validate spectral interpretation. This poses huge difficulties in mining large data-sets for modifications and in obtaining curated databases. Thus, the affirmation of Patterson in 2003 regarding proteomics and bioinformatics is still true: “the realization of the full potential of technological advances will require concurrent intensive efforts on the computational front” (120).

b) Chemical Proteomics for analysis of PTMs

As noted before, recent improvements in mass spectrometry and other proteomics technologies have extended the identification and quantitative analysis of protein expression in different pathophysiological conditions to include the study of PTMs. In recent years, there has been increasing interest in studying how these modifications regulate a wide range of protein functions, such as stability, cellular location as well as protein-protein interaction (75, 121, 122). Over the last few years, more than 900 different PTMs were reported and included in the Unimod database (http://www.unimod.org); however, when performing a bottom-up experiment, commonly only a few PTMs are detected, due to the proteomic complexity of biological samples. Methods to enrich modified proteins/peptides are then required to reduce sample complexity, increasing the number of identified PTMs. Thus, a range of methods for selective enrichment of modified proteins has been developed for the identification and quantification of PTMs such as phosphorylation, acetylation, glycosylation, sumoylation, ubiquitination, oxidation, nitration, and nitrosylation. Depending on the purpose of the enrichment technique, two main approaches have been used (figure 1): enrichment at the protein level, or enrichment at the peptide level after digestion of whole protein lysates with proteases. The following sections will review recent developments in PTM enrichment methods, such as those used for phosphorylation, acetylation, glycosylation and oxidation.

Protein/peptide enrichment strategies for PTM analysis

Enrichment strategies of several non-oxidative modifications have been well researched and applied, and provide a useful illustration of the principles and limitations that also underlie enrichment for oxidative PTMs.

Phosphorylation

Among all the identified PTMs, phosphorylation is probably the most extensively studied and at least 10,000 proteins have been described to undergo phosphorylation (123-125). It has been known for decades that this PTM plays an important role in cell cycle, apoptosis, metabolism, receptor function and stress responses, through the activation of signaling cascades (126-128). Since phosphorylated proteins/peptides constitute a minor percentage of the total proteome, enrichment techniques are considered an essential step for the MS-based phosphoproteome profiling. Phosphorylation is a dynamic modification, and care should be taken to avoid protein degradation and loss of phosphate groups due to the activity of proteases and phosphatases (129). There are several enrichment strategies for phosphoproteins and phosphopeptides. The most used are immobilized metal ion affinity chromatography (IMAC) (130), titanium dioxide (TiO₂) (131) (reviewed in (122)) and immunoaffinity chromatographies (reviewed in (132)). Nowadays, for phosphoproteome profiling the phosphopeptide enrichment is usually a preferred approach. The most used method for phosphopeptide enrichment is based on the affinity of phosphate groups to metal oxides, in particular TiO₂. This method is highly selective and sensitive for the analysis of phosphorylated peptides by MS. For example, using this approach, more than 3,000 distinct
phosphopeptides have been identified in different experiments (81, 128). Recently, several affinity resins consisting of ionic metals or metal oxides were investigated for their phosphopeptide enrichment capabilities (133). The authors have concluded that this procedure dramatically improve detection and sequencing of phosphopeptides, compared with analyses without enrichment. However, this approach is biased towards detecting monophosphorylated phosphopeptides, since multiphosphorylated peptides bind with high affinity to TiO$_2$ and may not elute (125, 134). In order to increase the identification of multiphosphorylated peptides, multidimensional enrichment approaches were developed, combining TiO$_2$ with a sequential elution from IMAC and hydrophilic interaction liquid chromatography (HILIC) separation procedures. This combined strategy, collectively abbreviated TiSH, has been used for the identification of more than 6,000 unique phosphopeptides in insulinoma cells (134). IMAC was initially developed for the separation of histidine-rich peptides and proteins from complex mixtures. The selection of the metal ion, such as Fe$^{3+}$ and Ga$^{2+}$, was exploited to enhance the specificity for oxygen-rich functional groups such as the phosphate anion. Nevertheless, despite allowing the enrichment of multiphosphorylated peptides, this methodology was shown to be less selective than TiO$_2$ (reviewed in (125)). Despite the methodological improvements in the selective isolation of phosphopeptides either with IMAC or TiO2, the MS analysis of phosphopeptides remains a challenge due to the unique and labile chemistry of phospho-Ser/Thr residues. Also, the presence of phosphorylated amino acid residues compromises the fragmentation efficiency in collision-induced dissociation (CID) causing the neutral loss of phosphate or phosphoric acid (m/z 98, 49, or 32.7 depending on whether the peptide is singly, doubly or triply charged) that precludes the localization of phosphosites (135, 136). Nevertheless, the presence of dehydroalanine or dehydro-2-amino butyric acid in the tandem mass spectra can be used as indicators of a serine or threonine phosphorylated peptide (137). Immunoadfinity chromatography enrichment, based on phospho-specific antibodies, continues to be largely used, in part due to the highly selective phospho-specific antibodies available that can be utilized for the enrichment of targeted phosphorylated proteins (138). Moreover, anti-pSer/pThr antibodies can be used, aiming to reveal novel kinase substrates (139). However, the rather low specificity of some of these antibodies, probably the exception of anti-phosphotyrosine, limits their general use in phosphoproteomics (140). Nevertheless, the development and availability of a broad range of specific anti-phosphoamino acid antibodies an in-depth profiling of phosphoproteome dynamics can be expected in the near future. Since each phosphorylation site results from a kinase reaction, this phosphorylation data can be used to study signaling pathways (141).

Glycosylation

Protein glycosylation is one of the most ubiquitous and complex PTMs, playing a key role in protein stability, folding and solubility, and modulating cellular processes such as cell adhesion and differentiation (142, 143). Alterations in glycosylation have been associated with different pathological conditions. However, the underlying interrelation between this PTM and physiological states is still poorly understood and intensive research aimed at glycoprotein/glycopeptide characterization has been conducted, mainly with the aid of MS methodologies (144-147). Glycosylation is one of the most abundant protein posttranslational modifications. To study protein glycosylation, enrichment strategies have also been developed for the study of this PTM. Protein glycosylation encompasses a myriad of different modifications, with variations arising from the size and nature of the polysaccharide to the characteristic site of glycan linkage. Hence, selection of the ideal enrichment methodology is highly dependent on the type of glycoproteins under study. For the
analysis of N-glycoproteins, most of the enrichment procedures are based on hydrazide chemistry, involving oxidation of the carbohydrates in glycoproteins for their subsequent immobilization onto a hydrazide-activated resin, followed by the release of glycopeptides with Peptide N-Glycosidase F (PNGase F) (148). Proteolysis, derivatization and MS analysis facilitate the identification of glycoproteins and the respective assignment of N-glycosylation sites (149). Nevertheless, this experimental approach does not provide information regarding glycan structure and degree of site occupancy due to the oxidative chemical coupling of the glycan to hydrazide (148). To overcome this limitation, a strategy based on mild periodate oxidation was developed for the characterization of sialic acid-containing glycopeptides (150). With this procedure it is possible to maintain the glycan structure stable with the exception of sialic acid moieties (151). For the characterization of O-glycoproteins, lectins have been largely exploited (152). Lectins are proteins or glycoproteins ubiquitously found in nature that bind reversibly to specific glycan structures. Most lectins interact exclusively with the terminal non-reducing position in an oligosaccharide, but some, such as Concanavalin A (ConA) and wheat germ agglutinin (WGA), bind to internal sugars (153, 154). Indeed, there is an array of different lectins commercially available and their selection depends on the structural motifs of the glycans to be targeted. For instance, wheat germ agglutinin (WGA) and Maackia amurensis (MA) have been used for enrichment in α2,3-linked sialic acid moieties in O- and in N-glycoproteins (155). ConA shows a broad spectrum affinity for mannosyl and glucosyl residues in glycans containing free hydroxyl groups at C3, C4 and C6 (156). Moreover, lectins can be immobilized in distinct solid supports (e.g. silica (157), agarose (158) or sepharose (159)), monolithic columns (160), and in microdimensional systems (146, 153, 156, 161, 162), including magnetic particles (142, 153). These magnetic beads coupled to lectins have been recently used for capturing glycoproteins in complex biological samples, taking advantage of their high surface area and high mobility in solution (142, 161). Immobilization or covalent capture of the target molecules seems promising, since it allows extensive washing without significant loss in sensitivity (161). Taken together, when the aim is the broad range analysis of protein glycosylation, lectin arrays seem to be the most suitable strategy to cover different types of glycan structures and to study their relationship with cellular dynamics (143, 163).

An alternative approach for glycoprotein enrichment uses boronic acid, taking advantage of the formation of cyclic boronate esters at high pH with saccharides such as mannose, galactose and glucose (159). In contrast to lectins, this enrichment methodology does not require a complex recognition motif, and allows the capture of both N- and O- as well as non-enzymatic glycopeptides (146). However, boronic acid-functionalized materials often display non-specific binding when complex samples with high amounts of non-glycosylated peptides are analyzed (164).

Acetylation

In recent years, protein acetylation has emerged as a fundamental PTM, rivaling with phosphorylation in its relevance to the regulation of biological processes (165). Proteins may become acetylated either at the α-amino group of the protein N-terminus or the ε-amino group of Lys residues. The binding of an acetyl group to these amines leads to their deprotonation, thus eliminating their characteristic positive charge. The reversible nature of Lys acetylation presents a key role in the interaction between proteins and other biomolecules, in particular with DNA, but also with other targets (165). The enrichment approaches for the study of the acetylome dynamics are based on immunoaffinity purification. For example, using immunoaffinity purification with anti-
acetyl-Lys antibodies, Mann and coworkers described the identification of 3600 acetylation sites in 1,750 human proteins and quantified acetylation changes, showing that lysine acetylation plays a role in the regulation of a diverse set of cellular functions (166). Using the same approach, more than 3,000 different acetylation sites in approximately 2,000 distinct proteins were identified in liver mitochondria and tentatively related with metabolic reprogramming (167). Combined fractional diagonal chromatography (COFRADIC) was specifically developed for the isolation and characterization of methionine-containing peptides (168). This technique can also be used to both qualitatively and quantitatively assess protein N\textsuperscript{α}-acetylation in whole proteomes (169).

**Ubiquitination**

Protein ubiquitination is known to regulate protein degradation, signal transduction, intracellular localization and DNA repair, depending on the protein and modification site targeted by this modification (170, 171). This PTM results from the covalent attachment of ubiquitin, a 76-residue polypeptide, to proteins at specific lysine residues. Tryptic digest of ubiquitinated proteins present a small di-glycine signature motif at the modification site (172), allowing ubiquitination to be detected and located during peptide analysis by MS/MS. For enrichment, immunoaffinity purification of ubiquitinated peptides was developed utilizing an anti-diglycyl-lysine monoclonal antibody (173). Adopting this approach, Mann and coworkers identified more than 10,000 ubiquitinated sites in human cells (170), highlighting the widespread nature of this modification in proteomes. Also, an immunoaffinity purification method utilizing exogenously tagged ubiquitin has been developed by Danielsen *et al.* (174).

As stated before, a range of methods for selective enrichment of modified proteins has been developed for the identification and quantification of PTMs. It is out of the scope of this review to discuss all of these methods, so we have kept out this discussion other more infrequently observed PTMs. For example, for protein S-acylation (palmitoylation), a reversible post-translational modification, methods including metabolic labeling of cells with alkynyl-palmitate and methods that protect free cysteines with thiol-reactive reagents, followed by removal of S-palmitoylation with NH\textsubscript{2}OH have been developed (reviewed in (175)). Recently, this later method was used to profile S-palmitoylated proteins in the murine RAW264.7 macrophage line (176).

**Detection and enrichment of oxidative post-translational modifications using proteomics**

Oxidative post-translational modifications of proteins are increasingly understood to play important roles in regulation of protein function and activity (177-179). Oxidation of proteins can take place at a variety of different residues in proteins, some of which are more susceptible than others to oxidation and the effect on the protein and the system depends on the exact nature and site of the modification (56, 180). The types of oxidations that occur can be divided essentially into reversible modifications, which are likely to be involved in regulatory processes, and irreversible modifications, which are unlikely to be involved directly in signaling. The residues most susceptible to oxidation are those containing sulphur, namely cysteine and methionine. Most of the modifications occurring at these residues are reversible and have been described to be involved in regulatory processes. Oxidation of other amino acids may have alternative biological effects through loss of function, and may have value as biomarkers of oxidative stress. Detecting these protein oxidative...
PTMs by mass spectrometry has proved to be a difficult task since these modifications are often of very low abundance and chemically unstable. Additionally, it may be difficult to ascertain whether the oxidation of highly reactive residues, such as methionine (181) or cysteine, represent artifacts upon sample handling, or true PTMs. Tomer and co-workers, have shown that tryptophan oxidation may occurs following to protein purification and isolation, particularly with the use of gel electrophoresis (182); and electrospray ionization has also been reported to induce oxidation of methionine, trytophan or tyrosine residues (183).

As stated earlier, for MS/MS identification of PTM’s, it is necessary that the mass shift detected in the precursor ions (peptide from tryptic digestion) is also observed in the fragment ions carrying the modified amino acid residue. Table 1 lists examples of mass shifts resulting from oxidation of amino acid residues, compiled from the UNIMOD database (http://www.unimod.org), a repository for protein modifications. A comprehensive review compiling the list of amino acid modifications induced by oxidative stress and detected by mass spectrometry has been published by Möller and co-workers [188].

<table>
<thead>
<tr>
<th>Modification</th>
<th>Discription</th>
<th>Δm (monoisotopic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbonyl</td>
<td>Formation of carbonyl moiety (Arg, Gln, Glu, Ile, Leu, Lys, Val)</td>
<td>13.9793</td>
</tr>
<tr>
<td>Oxidation</td>
<td>Oxygen addition and hydroxylation (Lys, Arg, Cys, Met, Tyr, His, Pro, Trp, Phe, Asp, Asn)</td>
<td>15.9949</td>
</tr>
<tr>
<td>Dioxidation</td>
<td>Addition of two oxygen atoms (Lys, Arg, Cys, Met, Tyr, Pro, Trp, Phe)</td>
<td>31.9898</td>
</tr>
<tr>
<td>Carbamylation</td>
<td>Isocyanate reaction with amino groups (Lys, Arg, Cys, Met)</td>
<td>43.0058</td>
</tr>
<tr>
<td>Decarboxylation (Asp, Glu)</td>
<td>Arginine oxidation to glutamic semialdehyde (Arg, Asn, Gln)</td>
<td>0.9840</td>
</tr>
<tr>
<td>Amino (Tyr)</td>
<td>Tyrosine oxidation to 2-aminotyrosine</td>
<td>15.0108</td>
</tr>
<tr>
<td>Arg → GluSA</td>
<td>Arginine oxidation to glutamic semialdehyde</td>
<td>-43.0534</td>
</tr>
<tr>
<td>Cys → Oxoalanine</td>
<td>Cysteine oxidation to o xoalanine</td>
<td>-17.9928</td>
</tr>
<tr>
<td>Cysteinylation (Cys)</td>
<td>Addition of free cysteine to form a disulfide (Cys)</td>
<td>+119.1423</td>
</tr>
<tr>
<td>Glutathionylation (Cys)</td>
<td>Addition of glutathione to form a disulfide (Cys)</td>
<td>+305.3076</td>
</tr>
<tr>
<td>S-nitrosylation (Cys)</td>
<td>Thiol adduct with nitric oxide</td>
<td>28.9902</td>
</tr>
<tr>
<td>Trioxidation (Cys)</td>
<td>Cysteine oxidation to cysteic acid</td>
<td>47.9847</td>
</tr>
<tr>
<td>His → Asn</td>
<td>Histidine oxidation to asparagine</td>
<td>-23.0159</td>
</tr>
<tr>
<td>His → Asp</td>
<td>Histidine oxidation to aspartic acid</td>
<td>-22.0319</td>
</tr>
<tr>
<td>His → Aspartylurea</td>
<td>Histidine conversion to aspartylurea</td>
<td>-10.0320</td>
</tr>
<tr>
<td>His → Formylasparagine</td>
<td>Histidine oxidation to formylasparagine</td>
<td>4.9790</td>
</tr>
<tr>
<td>Lys → Allysine</td>
<td>Lysine oxidation to aminoadipic semialdehyde</td>
<td>-1.0316</td>
</tr>
<tr>
<td>Lys → Aminoadipic acid</td>
<td>Lysine oxidation to α-aminoadipic acid</td>
<td>14.9632</td>
</tr>
<tr>
<td>Met → Aspartate semi aldehyde</td>
<td>Methionine conversion to aspartate semialdehyde</td>
<td>-32.0085</td>
</tr>
<tr>
<td>Met → Homocysteic acid</td>
<td>Methionine oxidation to homocysteic acid</td>
<td>33.9691</td>
</tr>
<tr>
<td>Pro → Pyrrolidine</td>
<td>Proline oxidation to pyrrolidine</td>
<td>-27.9949</td>
</tr>
<tr>
<td>Pro → Pyrrolidinone</td>
<td>Proline oxidation to pyrrolidinone</td>
<td>-30.0105</td>
</tr>
<tr>
<td>Pro → Pyro-Glu</td>
<td>Proline oxidation to pyroglutamic acid</td>
<td>13.9792</td>
</tr>
<tr>
<td>Trp → Hydroxykynurenin</td>
<td>Tryptofan oxidation to hydroxykynurenin</td>
<td>19.9898</td>
</tr>
</tbody>
</table>
Oxidation of sulphur containing residues:

Methionine is often present on the surface of proteins, and was suggested in the 1980s to act as a scavenger of oxidants in order to protect other residues in the protein from oxidative damage (184). However, more recently evidence for a role of methionine residues in the redox regulation of enzyme activity has been found. A well-established case is that of calmodulin, a calcium regulatory protein, where the oxidation of two methionines has been reported to cause structural changes, altering its interaction with target proteins (185, 186). The role of cysteine residues in redox regulation is determined by the chemistry of its thiol function. Reversible oxidation of cysteine thiolates may yield intra-molecular disulfides, protein mixed disulfides with low molecular weight compounds (eg. glutathione), S-nitrosothiols, or sulfenic acid. These reversible oxidations of sulphur-containing residues are thought to be of substantial importance in a variety of signalling pathways, in a manner analogous to and complementary with protein phosphorylation (56). Furthermore, cysteine can be irreversibly oxidized from the thiol form to sulfinim and sulfonic acids by addition of 2 or 3 oxygen atoms respectively. The initial oxidation to sulfenic acid (addition of 1 oxygen atom) is most favored for cysteines with a low pKa that exist in the thiolate form (187), and has been reported to play an important role in regulation of the activity of several signalling proteins (178). The sulfenic acid can react intra- or inter-molecularly with another thiol group to generate a disulfide bond, which can subsequently be reduced by thioredoxin to regenerate the original thiol. S-nitrosylation of cysteine is also a reversible modification and the thiol group may be regenerated by the action of glutathione (188). There is a large number of redox regulated proteins, but one of the best described examples is the protein tyrosine phosphatase PTP1B, where it is known that oxidation of the catalytic cysteine residue to a sulfenic acid inactivates the enzyme (189).

Mass spectrometry takes advantage of the mass changes occurring upon thiol oxidation in order to identify oxidized cysteines (table 1). Additionally, during collision induced dissociation (CID) the occurrence of characteristic neutral losses is regarded as a signature for these PTMs. Sulfenic acid modified peptides show a loss of 50 Da (-H\(_2\)SO) (83), while sulfinic and sulfonic acid modified peptides show losses of 66 Da (-H\(_2\)SO\(_2\)) and 82 Da (-H\(_2\)SO\(_3\)) respectively (85). Similarly, fragmentation of methionine sulfoxide-containing peptides shows a characteristic loss of 64 Da, corresponding to the dissociation of a methanesulfenic acid group (CH\(_3\)SOH) (190). Besides the direct detection of thiol modifications in complex protein mixtures, several methods have been developed for detecting and selecting modified thiols (reviewed in (191, 192)). These include the enrichment of affinity-tagged derivatives and biotin chromatography (193, 194) or Ni chromatography (195). A major breakthrough occurred with the development of a biotin-switch method for the affinity selection of S-nitrosylated peptides, by Jaffrey and Snyder (196). The principle behind this method involves alkylation of free thiols with methylmethanethiosulfonate (MMTS), followed by the selective reduction of the nitrosylated thiols to give a free thiol that is then available to be labeled by biotin. Since then, similar methods for detecting sulfenated (RSOH) and glutathionylated proteins have also been developed (197, 198). These methods utilize arsenite and glutaredoxin as specific reductants before labeling the thiol with the biotin tag. A quantitative enrichment method for oxidative thiol modifications has also
been developed, which uses isotope coded affinity tag (ICAT) technology combined with a differential thiol trapping technique (OxICAT) (199).

Oxidation of non-sulphur containing residues:

The oxidation of non-sulphur containing amino acid residues in proteins is mainly determined by the nature of their side-chain and their rates of reactions towards free radicals (200). The residues most susceptible to oxidation include the aromatic residues tyrosine, tryptophan, phenylalanine and histidine. A typical oxidative modification of these residues is hydroxylation, although this may lead to further oxidation and degradation, such as the breakdown of hydroxylated tryptophan to kynurenine and N-formylkynurenine (201) (table 1). Lysine and arginine can undergo oxidative deamination to form carbonyl groups, but these can also be formed on serine and threonine residues (180).

Protein oxidation by reactive oxygen species (ROS) may also result in important biological effects. The most obvious and common effect of oxidation is the loss of normal activity or function. This is most likely to occur if the residue(s) affected is within or adjacent to the catalytic or substrate binding sites, in the case of an enzyme, or the ligand binding domain in the case of receptors. It appears that individual oxidations on the surface of the protein often have little effect on its conformation and therefore its activity (202), and substantial modification is needed to produce a noticeable change (203). A further physiological and frequently pathological aspect of protein oxidation is the potential of more severe oxidation to cause partial unfolding of the native structure, resulting in exposure of hydrophobic regions of the protein (204, 205). This tends to lead to aggregation of protein molecules, which can be a serious problem as the proteasome may no longer be able to remove them, and they accumulate within the cell or tissue, often with detrimental effects (204, 206). A similar outcome can arise from less severe oxidation if cross-linking of proteins occurs, such as tyrosine cross-linking (207, 208).

Mass spectrometry has played a major role in the study of protein oxidative modifications. Since the seminal work by Finch et al. (209) on the identification of oxidative modifications in serum albumin, in vitro oxidative stress systems have been successfully employed to determine the mechanisms of protein oxidation (210-215). These contributions are considered fundamental both for identifying oxidative modifications that can be considered signatures of ROS activity in biological samples and for the development of the analytical methods necessary for characterizing protein oxidation in vivo. Additionally, several studies with model peptides have proved to be fundamental to determine the susceptibility of amino acid residues to oxidation and to understand the chemical rules governing protein oxidation (216-218). Data describing protein oxidation in vivo is relatively scarce, and most studies that have attempted to identify potential oxidative stress markers have not been able to associate the PTMs with significant biological effects. The study of oxidative PTMs in diabetes mellitus has been undertaken in samples including human saliva (219), rat plasma (220) and mouse liver mitochondria (221). Protein oxidation has also been investigated in association to other pathological conditions such as breast cancer (222), epilepsies (223) or spinal cord injury (224).

Much work on identification and isolation of oxidized proteins has relied on the detection of protein carbonyls; these are one of the most commonly measured indicators of protein oxidation, owing to the relative simplicity and cheapness of the available assay (180, 225, 226). Carbonylated peptides have been isolated from complex protein mixtures through their derivatization with Girard
P reagent (1-(2-hydrazino-2-oxoethyl)pyridinium chloride), or biotin-hydrazide, followed by protein digestion and peptide separation by strong cation exchange (SCX) or avidin chromatography (227). DNPH-labeled carbonylated peptides have also been detecte using a precursor ion-like scan in negative ion mode (228). Applying the biotin-hydrazide derivatization and selection method, Regnier and co-workers were able to identify oxidized proteins in the plasma of breast cancer patients (222) and show that the carbonylation levels of very significant number of proteins changed 1.5 fold or more (95). This result demonstrates the potential of oxidative PTMs as biomarkers of elevated oxidative stress conditions in disease states.

Protein modification by cross-reaction with carbohydrate or lipid oxidation products:

The occurrence of carbonyl groups on proteins can result either from the direct oxidation of side chains or by cross-reaction with a secondary product of lipid or carbohydrate oxidation. During lipid peroxidation and during the glycoxidation of carbohydrates, an array of reactive carbonyl species (RCS) is formed, which are precursors of advanced glycation end products (AGEs) and advanced lipid peroxidation end products (ALEs) (229, 230). (231). These RCS which, among others, include glyoxal, methylglyoxal, malondialdehyde, acrolein, 4-hydroxy-2-nonenal (HNE), and 4-oxo-2-nonenal (ONE), bind covalently to proteins via Michael addition or Schiff base formation, involving reaction with nucleophilic groups in the side chains (232). Schiff base formation involves the reaction of carbonyl groups with the amine of lysine residues, so protein carbonyls only arise if the RCS contains 2 carbonyl groups. In contrast, Michael adducts are formed by reaction of electrophilic α,β-unsaturated moieties predominantly with side chains of cysteine, histidine or lysine, leaving the carbonyl group of the RCS available for further reactions. For instance, the β-cleavage of hydroperoxides derived from the oxidation of polyunsaturated fatty acids generates 4-hydroxy-2-nonenal (HNE), a well-known reactive aldehyde which cross-reacts proteins causing their irreversible inactivation or degradation (reviewed in (233)). Mildly HNE-modified proteins are rapidly degraded by the proteasome, but extensive modification causes protein cross-linking which in turn impairs proteasome function compromising cellular protein turnover (234). Using animal models of ethanol-induced oxidative stress, it was shown that HNE consistently modified heat shock proteins 72 and 90 (HSP72 and HSP90) (235, 236), suggesting a role of HNE in compromising protein homeostasis.

HNE-protein adducts are commonly enriched by immunoaffinity chromatography (237) or by a solid-phase hydrazide strategy (238). Mass spectrometry analysis of these modified proteins has been carried out under several conditions (reviewed in (232) and (239)). Most of the formed adducts are chemically reversible and only those which prove to be stable under the experimental conditions adopted will be detected. A common analytical approach includes a stabilization step by reduction with borohydride (NaBH₄) prior to protein digestion (240). MS studies with non-reduced peptide adducts have demonstrated that the major product ions upon tandem MS fragmentation result from the neutral loss of the aldehyde residues, dificulting the assignment of the modification site (241, 242). Ultimately, this chemical property may provide useful information through neutral loss dependent ion selection for MS/MS peptide identification or data-dependent neutral-loss driven MS3 acquisition(88).

c) Quantification strategies for PTMs analysis
In order to determine the functional role of modified proteins in biological systems, efforts have been made to compare their abundance quantitatively under different physiological conditions. However, the quantification of PTMs not only encompasses the determination of protein abundance, but also requires the determination of the occupancy levels at the modification site (243, 244). Several MS-based quantitation standard methods for quantification of peptides can be used to determine PTMs abundance. These may be grouped into either stable isotopic labeling or label-free workflows.

Stable isotope incorporation can be achieved through metabolic labeling in cell culture (in vivo) or by chemical labeling in a post-metabolic context (in vitro). For the metabolic labeling of proteins, two strategies are widely used: 15N enrichment of cell culture medium (245); addition of stable isotope encoded amino acids such as 3D-Leu or 13C6-Arg to the cell culture medium, an approach known as stable isotope labeling with amino acids in cell culture (SILAC)(246). These approaches allow the quantitative comparison of two to five different samples (247). For the post-metabolic labeling strategy, the chemical derivatization of the peptide N-terminal or C-terminal is generally performed after protein extraction and proteolytic digestion. In the case of isobaric tags for relative and absolute quantitation (iTRAQ) (248) or tandem mass tag (TMT) (249), the primary amines of proteolytic peptides are tagged via N-hydroxy-succinimide chemistry, maintaining the same physicochemical properties in identical peptides from different samples. The tags all have identical masses, but generate different fragmentation products, e.g. at m/z 114, 115, 116 and 117. Therefore identical peptides obtained from different samples can be selected for fragmentation as a single precursor ion, but generated distinct reporter ions in the MS/MS spectrum, allowing the relative levels of the peptide (and hence the protein) from different samples to be compared. This approach allows the simultaneous analysis of four to eight distinct samples, depending on the reagents used (250). The interested reader may refer to (251) for an updated review of iTRAQ methods and applications. All these isotope labeling methodologies have been successfully employed in the quantification of PTMs, for example, in the characterization of grapevine response to infection by simultaneous monitoring protein phosphorylation, acetylation and glycosylation (252). Also, different variations of this methodology have been applied recently to detect redox changes in thiol proteins(253-255) Nevertheless, these approaches might be impractical in some cases, such as peptide-centric workflows that target PTM containing peptides or immunoaffinity enrichment procedure which are incompatible with the chemical tags (256).

Label-free quantification methods are a suitable alternative to metabolic and post-metabolic labeling. This quantification methods do not rely on any kind of peptide or protein labeling: quantification results from the computational analysis of spectral counting (defined as the sum of MS to MS/MS transitions) or MS ion intensity using different algorithmic approaches (reviewed in (257)). Whereas spectral counting is considered a more appropriate approach for relative protein quantification over a small dynamic range, methods based on ion abundance obtained from MS scans seem to be more compatible with the quantification of PTM-modified peptides (256). Since the chemical derivatization of peptides or proteins is not required, label-free methodologies have a wide range of applications in PTM screening, for example in epigenetic studies based on histone modifications (258, 259).

Report on 10 Years of PTM Research (2003-2012)
Until now, there have been 72320 post-translational modifications reported experimentally in the UniProt Knowledgebase (UniProtKB), of which 49062 were phosphorylations, 5736 N-linked glycosylations and 5164 acetylations (252, 260). Oxidative PTMs are included within, but do not exclusively comprise, the remaining 12,358 modifications, so it can be seen that there has been comparatively less focus on them in the proteomics field. To the following sections provide a comprehensive analysis of bibliography in the PTM field, reflecting the scientific achievements and the recent technical developments in the detection of PTMs. In order to ascertain the state of the art in mass spectrometry and proteomics methodologies for the study of PTMs, we used two approaches: analysis of the PTM data introduced in the Universal Protein Resource (UniProt) (261) and analysis of the literature published in the last three years (2010, 2011 and 2012).

a. Results from UniProt

The UniProt knowledgebase (UniProtKB) (http://www.uniprot.org/help/uniprotkb) is used to access high quality annotated non-redundant sequence and functional information on proteins, and is nowadays one of the most important sources of proteomic information. It contains manually annotated information that was extracted from literature and curated using computational analysis (262).

The number of proteins annotated as being post translationally modified were 22538 in January of 2013. As it can be seen in figure 2, this number increased exponentially until 2008, when 5714 proteins were entered, showing that a high degree of PTM processing and analysis by the scientific community took place. However, for unknown reasons to the authors, since then the number has dramatically decreased to 55 in 2012 (figure 2). It is interesting to note that few proteins in UniProt are annotated as oxidized (only 38 annotations), clearly indicating a relatively small effort and generation of data in this specific area.

![Figure 2](image)

**Figure 2** – Yearly distribution of the number of proteins found to be post-translationally modified using mass spectrometry, as annotated in UniProt.

The first article cited in UniProt database corresponding to the identification of a PTM was published in 1978 (263). Since then, 1275 articles have been cited in UniProtKB reporting the identification of PTMs by mass spectrometry. Of these, 19 reported PTMs detected using a shotgun approach and the remaining using a focused approach. The total number of annotated articles in
UniProt was 1275 (122 using a large scale approach and 1153 using a focused approach). Analysis of the number of times that references reporting the use of mass spectrometry have been used for citing PTMs in the UniProtKB showed that most of annotated PTMs in UniProt were from large scale studies. In fact, the maximum number of reference entries was in 2008, with 10146 citations from large scale studies and 1278 from focused research studies. Since in 2008 21 articles reporting large scale studies were cited, this implies that most of the annotated proteins and PTMs were in fact the result from a few large scale studies.

b. Three Years of PTM identification and characterization by Mass Spectrometry (2010-2012)

We have collected information regarding the chemical nature of the identified PTMs, the experimental approach, the mass spectrometry instrument, the data analysis software and protein database selected by the authors, and the nature of the biological sample under study (supplementary table 1). Figure 3 shows the number of research papers published in each year, highlighting the number of studies in which a shotgun approach has been adopted. Despite the slightly higher number of studies published in 2010, when compared with the two subsequent years, we did not find any apparent difference between the three years. Being so, we believe that the chosen time period may be generically treated as homogenously representing the most up-to-data research in PTM characterization by mass spectrometry. The vast majority of the research works herein analyzed were found to deal with the characterization of protein phosphorylation (P), glycosylation (G), methylation (Met), acetylation (Ac) or oxidation (O) (Figure 3). Oxidation refers to the addition of a single oxygen atom to the structure, and therefore corresponds to a specific subset of possible oxidative modifications. Some other oxidative changes, such as nitration (5 articles), nitrosylation (6 articles), cysteinylation (5 articles) and glutathionylation (3 articles) were also reported, as shown in supplementary table 1.

Figure 3: Number of research articles reporting the characterization by mass spectrometry of protein post-translational modifications (PTMs) occurring in biological samples in the years 2010, 2011, and 2012. The graph shows the number of publications per year and the number of publications reporting the identification of phosphorylations (P), glycosylations (G), methylations (Met), acetylations (Ac), oxidations (O) or an array of different modifications (Other). Distinction is made
between the articles describing PTM identification through a large scale or shotgun experimental approach and a focused approach.

PTM enrichment

The significance of enrichment procedures, increasing the relative abundance of PTM-bearing peptides, is evident from supplementary table 1. If one focuses on the median value for the number of PTMs reported in the communications herein considered, it can be seen that utilizing enrichment procedures doubles the number of identifications. Enrichment procedures are preferentially adopted as part of a shotgun approach. From the 88 studies reporting on the use of PTM enrichment, 65 correspond to large scale proteome investigations. A third (22) of such studies has utilized phosphopeptide enrichment methodologies, the most popular being affinity to TiO$_2$ (employed 21 times), followed by other IMAC (8) and electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) (4). The median number of phosphosites reported in shotgun studies is 126 (mean = 1231). However, when we only consider studies where phosphopeptide enrichment was employed, this value rises to 306 (mean = 1693). The most popular method for enrichment in glycosylated peptides has been affinity selection with specific lectins (used in 11 studies). Enrichment procedures for other types of modifications, including methylation, acetylation, different forms of protein acylation and protein ubiquitination have been based on the immunoaffinity selection or immunoprecipitation of peptides containing the modification of choice (21 reports). The specific derivatization of the modification site with a biotin tag, subsequently selected by biotin-avidin binding, has been reported 8 times. During the time period under study we could only find the works published by Regnier and co-workers describing sample enrichment in carbonylated proteins. Selection of carbonylated proteins was achieved through the derivatization of the carbonyl moiety with biotin hydrazine followed by avidin affinity chromatography (220, 264). Despite these recent efforts, in the articles herein compiled, the number of oxidative modifications and oxidized proteins identified is generally small when compared to the numbers available for other modifications such as phosphorylation or acetylation. This is undoubtedly due to the difficulty in detecting oxidative PTMs, which arise from the variety of oxidation products, their low relative abundance and mainly the labile nature of some of these PTMs under tandem MS analysis conditions. The correct choice of instrumentation and particularly the application of the various fragmentation modes in tandem MS may prove fundamental to overcoming these difficulties. ETD, for example, can constitute a suitable alternative to the most commonly used CID fragmentation. ETD tends to produce extensive peptide backbone fragmentation, improving sequence information, while preserving labile amino acid side chain modifications.

Mass spectrometry approaches to PTM detection

In the last three years, almost all studies reporting the identification of PTMs by MS have utilized a similar analytical setup, consisting of on-line nano-LC-MS/MS. In the case of MALDI ionization sources, the chromatographic separation was carried out off-line from the MS instrument. We could only identify 6 studies in the period between 2010 and 2012 which had not utilized an LC-MS setup: these consisted of 3 top-down research articles and 3 focused analyses. The major instrumental setup differences result from the MS instrument adopted for the study. Figure 4A, represents the
frequency of use in PTM identification of the various mass spectrometers. The instruments were designated by their most distinctive feature, the mass analyzer. 28 studies used more than one instrument for PTM characterization and the pie chart total reflects this fact. Except for the instruments included in the group designated as TOF/TOF, all the other instruments utilize an ESI source, allowing complete interfacing with the liquid chromatography system. In the time period under analysis, Orbitrap instruments became the most popular, having been utilized in more than 40% (95) of the research projects on PTM characterization by MS. QTOF instruments were used in 49 studies and low resolution ion traps were employed 45 times. The utilization of FT-ICR instruments was reported 31 times, 11 of which were in top-down studies. However, there were only 4 papers in 2012 that describe the utilization of this instrument. Quadrupole-ion trap (QTRAP) and triple quadrupole (QqQ) instruments were used more sparsely and mainly with the purpose of identifying peptide modifications by targeted proteomics. MALDI sources were still found to be widely used (37). Most MALDI instruments have a TOF/TOF mass analyzer and only 5 studies employed MALDI-QTOF-MS instruments. Apart from the FT-ICR instruments, no significant differences in the frequency of use of the various instruments were found, either in the different years or in different experimental approaches. Furthermore, we have not been able to establish any clear advantage or preference in the utilization of a given instrument type for the study of a particular modification. Contrary to the predominance in the use of Orbitrap instruments for the characterization of other PTMs, there seems not be a clear trend in instrumentation preference for detection of protein oxidative modifications. 12 articles describe the utilization of Orbitrap instruments, while the use of other ion trap instruments has been described for 9 times. QTOF and MALDI-TOF/TOF-MS instrumentation has been employed 12 and 9 times, respectively. However, targeted approaches for detection of oxidative (or other) modifications can only be carried out in QQQ or QQLIT instruments. Nevertheless, precursor-like analysis can be done with high resolution, rapid scanning instruments.

Figure 4: Meta-analysis of the mass spectrometry instruments and data analysis procedures employed in PTM identification in the years 2010 to 2012. A) Chart describing the number of research articles describing the utilization of a given mass spectrometer instrument for PTM identification. Orbitrap, QTOF, IT, FT-ICR, QTRAP and QqQ describe instruments with the referred mass analyzer using an electrospray ionization (ESI) source. TOF/TOF instruments are utilized with a
matrix assisted laser/desorption ionization (MALDI). 5 studies were identified were a MALDI-QTOF-MS type instrument was used. B) Frequency of use of different peptide or protein quantification approaches. a) Label Free designates a quantification method based on chromatographic peak area integration, while other label free approaches are specified. b) Other than iTRAQ stable isotope labeling.

The effective assignment of a PTM site always requires the use of tandem mass spectrometry. CID is still the most common fragmentation method, with use of alternative fragmentation methods such as HCD, ETD and ECD for the characterization of PTMs in biological samples still appearing rarely. Altogether, the use of CID fragmentation has been reported in 216 of the research works herein analyzed, while only 46 publications described the utilization of fragmentation modes other than CID. In most of these studies (36 studies), particularly in bottom up approaches, the alternative fragmentation methods have been used as complementary to CID. Specifically, HCD was utilized in 18 studies, while ETD was used 23 times, both with Obitrap and IT instruments (supplementary table 1). ECD was used for 8 times with FT-ICR instruments, always in top-down experiments.

As yet, there are few studies where quantitation of PTMs has been attempted or successfully achieved. Only 66 of the curated research articles describe the quantification or semi-quantification of PTM or modified protein abundance from their datasets. 31 of these studies adopted a “focused” approach, while 35 were whole proteome investigations. Label-free methodologies lead the way for the quantification of PTMs. Spectral counting, chromatogram peak area and targeted approaches such as SRM and MRM have been utilized in 45 research works, while stable isotope labeling methods (iTRAQ, ICAT) and metabolic labeling (SILAC) have been used more rarely (figure 4B). The major difficulties concerning the mass spectrometry quantification PTMs arise from problems of sequence coverage, so that all modifications can be detected, difficulties of determining site occupancy, at least by bottom-up methods, and the labile nature of some modifications. In oxidized proteins, the multitude of possible oxidation products for a single amino acid residue poses additional problems in determining site occupancy. Additionally, in biological samples, due to repair or degradation processes, oxidized proteins tend to exist in low abundance, increasing the difficulty for their quantification in complex protein mixtures.

Conclusion
IMS based proteomics has emerged as a powerful tool in screening and quantification of PTMs. Although the current technology is unable to offer complete picture of the modified proteome, the recent advances in MS instrumentation and experimental approaches provide the means to explore the inter-relationship between the most prominent PTMs and to establish their function within cellular systems. Nevertheless, it is expected that the use of alternative fragmentation methods and the development of robust quantitative methods and effective enrichment methods (as those existing for phosphorylation) can change this scenario. Nowadays, the simultaneous large scale proteomic identification of different PTMs should be envisaged, if the global physiological role of
these modifications is to be assessed. Also, the current number of studies dedicated to the study of
PTMs in the pathophysiological conditions is scarce and it should constitute one of the primary goals
of this field of research in the near future. To fulfill this objective there is an urgent need to organize
and curate all the available information on PTMs. One of the more significant findings to emerge
from this review is the importance of a centralized complete database of all PTM assignments. Also,
considerably more work will be needed in bioinformatics in order to automatically validate the PTMs
assignments, particularly from large scale proteomic studies. Our analysis also shows that oxPTMs is
a field that needs to grow rapidly as application studies are very scarce in this field. In conclusion,
although the actual body of PTM information is large, and expected to increase exponentially in the
near future, there is an urgent need for finding platforms of PTM information integration.

_________________________________________________________________

Thanks are due to Fundação para a Ciência e a Tecnologia (FCT, Portugal), European Union, QREN,
FEDER and COMPETE for funding the QOPNA research unit (project PEst-C/QUI/UI0062/2011); the
Portuguese National MS Network, also supported by funds from FCT and to the project "New
Strategies Applied to Neuropathological Disorders", co-funded by QREN and EU, and COST Action
COST CM1001. AS would like to thank FCT for the Post Ph.D. grant (SFRH/BPD/64296/2009). CMS
would like to thank the Proxomics project (EP/I017887/1) for support.

References

1. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. The Journal of
2. Klose J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse
tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik.
1975;26(3):231-43. Epub 1975/01/01.
3. Scheele GA. Two-dimensional gel analysis of soluble proteins. Charaterization of guinea pig
1975/07/25.
4. Anderson L, Anderson NG. High resolution two-dimensional electrophoresis of human
6. Bravo R, Celis JE. Search for Differential Polypeptide-Synthesis Throughout the Cell-Cycle of
7. Rabilloud T, Chevallet M, Luche S, Lelong C. Two-dimensional gel electrophoresis in
8. Wilkins MR, Pasquale C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G,
Humphery-Smith I, Williams KL, Hochstrasser DF. From proteins to proteomes: large scale protein
identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology (N Y).


47. Hardman M, Makarov AA. Interfacing the orbitrap mass analyzer to an electrospray ion source. Analytical Chemistry. 2003;75(7):1699-705.


Supplementary Material

Supplementary Reference list 1: List of research articles published in the years 2010, 2011 and 2012 identifying protein post-translational modifications (PTMs) by mass spectrometry.

Supplementary Table 1: List of research articles published in the years 2010, 2011 and 2012 utilizing mass spectrometry as a tool for the characterization and identification of protein post-translation modifications (PTMs). The research articles published in these last three years were gathered utilizing Scopus (www.scopus.com) and the search was limited to the subject areas of “Life Sciences” and “Physical Sciences” with the following parameters: data range: 2010 to 2012; Document type: “All”; Search field type: “Article Title, Abstract, Keywords”. The search terms “Protein”, “Mass Spectrometry” and “Post-translational modification” were utilized in different search fields connected by the Boolean operator “AND”. Additional searches were performed, substituting the latter search term by other usual forms to designate PTMs: “Post-translational-modification”, “post translational modification”, “posttranslational modification”, “postranslational modification” and “PTM”. Finally, the “Search history” tool of Scopus was used to combine the various independent searches. The list of references was imported into the reference managing software “EndNote X5” (http://endnote.com/) and consisted of 970 entries, after excluding redundant references and works written in a language other than English. The articles retrieved by Scopus were curated manually and the following articles were excluded: reviews, papers describing the development of bioinformatic algorithms and software, and research articles describing the characterization of PTMs in peptide standards, protein standards, recombinant proteins or overexpressed proteins. At the end, a group of 226 references (Supplementary reference list 1) were selected, which was considered to represent the most recent advances in the characterization of protein PTMs in biological samples.