Interpreting the protein language using proteomics

Ole N. Jensen

Abstract | Post-translational modifications define the functional and structural plasticity of proteins in archaea, prokaryotes and eukaryotes. Multi-site protein modification modulates protein activity and macromolecular interactions and is involved in a range of fundamental molecular processes. Combining state-of-the-art technologies in molecular cell biology, protein mass spectrometry and bioinformatics, it is now feasible to discover and study the structural and functional roles of distinct protein post-translational modifications.

ChIP-on-chip
A chromatin immunoprecipitation (ChIP) experiment that uses protein–DNA crosslinking, specific antibodies and cDNA microarrays ('chips') for large-scale studies of DNA-binding proteins, including the analysis of post-translational modifications.

Protein Research Group, Department of Biochemistry and Molecular Biology, University of Southern Denmark, DK-5230 Odense M, Denmark.
e-mail: jenseno@bmb.sdu.dk
doi:10.1038/nrm1939
Published online 17 May 2006

Living cells rely on an intricate interplay between thousands of different biomolecules that maintain cellular integrity and morphology and execute many biological functions. Whereas cognate genes encode the basic biological functions of proteins, the real-time dynamics and regulation of protein structure and function are generally carried out by specific post-translational modifications (PTMs) of proteins such as phosphorylation, glycosylation and acylation. During recent years, protein PTMs have attracted attention in the biological and biomedical research communities, and it is now clear that most proteins in species from archaea to humans carry various site-specific covalent modifications[1–5] (FIG. 1).

Traditionally, each type of PTM — for example, phosphorylation — has been studied at a specific amino-acid residue or at a small number of amino-acid residues in an individual protein, and this approach is based on the idea that one modification site corresponds to one regulatory function. However, it is now evident that modified sites in proteins can not only mediate individual functions, but can also function together to fine-tune molecular interactions and to modulate overall protein activity and stability [5]. Multi-site phosphorylation is now recognized as an important mechanism for the regulation of protein localization and functional activity and for the attenuation of protein–protein and protein–ligand interactions[6], although the molecular details and mechanism of regulation remain elusive[7].

A prime example of complex multi-site modification is the 'histone code', which regulates the structure and function of nucleosomes in chromatin. The histone code is dynamically altered by a battery of enzymes that (reversibly) methylate, acetylate, phosphorylate and ubiquitylate distinct amino-acid residues in the core histones (FIG. 1), and thereby modulate their interactions with nucleic acids and regulatory protein factors during DNA replication, transcription and repair[8,9]. Using chromatin immunoprecipitation (ChIP) with modification-specific antibodies combined with DNA microarray analysis (ChIP-on-chip), it is possible to follow histone modifications at individual genes and promoters at the genomics level[10]. At the individual histone level, the interplay between neighbouring PTMs on histone H3 was discovered by using mass spectrometry (MS) to detect distinct trimethylation and phosphorylation events that adjust protein–protein interactions[11]. Many PTMs of histones have been determined by MS, and new details of the histone code are regularly reported[12]. The ability to interfere with the histone code by using designer inhibitory drugs that are aimed at histone deacetylases is considered to be a promising approach to block gene activation and transcription in transformed cells as a way to eradicate tumours[13]. PTMs therefore have a fundamental role in cell biology and are recognized as important targets in molecular medicine and pharmacology.

The fact that proteins carry many modifications increases the molecular heterogeneity and diversity of gene products. As a result of allelic variations, mRNA splicing and PTMs, the human genome can generate hundreds of thousands of different gene products from an estimated 22,500 open reading frames. However, only a distinct subset of these gene products is present in a given cell type under normal physiological conditions.

The question as to whether there is an interdependence or functional cooperation between the various PTMs of a protein remains to be systematically investigated by proteomics approaches. This is the core of the 'protein language' — that is, the types and patterns of PTMs that are present on a protein at a given time and under given conditions. The pattern of PTMs on proteins constitutes a molecular code that dictates protein conformation,

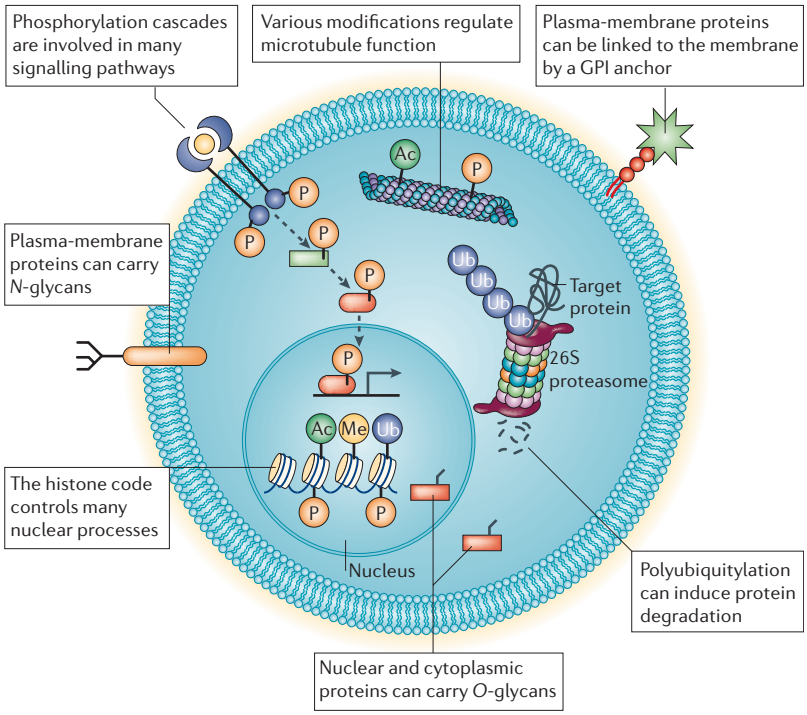


Figure 1 | **Cellular post-translational modifications.** This schematic figure shows the location and role of a selection of some of the most important of more than 200 types of post-translational modification (PTM). PTMs are found on all types of protein, from nuclear transcription factors to metabolic enzymes, structural proteins and plasma-membrane receptors. PTMs affect the physicochemical properties of proteins, which provides a mechanism for the dynamic regulation of molecular self-assembly and catalytic processes through the reversible molecular recognition of proteins, nucleic acids, metabolites, carbohydrates and phospholipids. Ac, acetyl group; GPI, glycosyl-phosphatidylinositol; Me, methyl group; P, phosphoryl group; Ub, ubiquitin.

cellular location, macromolecular interactions and activities, depending on the cell type, the tissue and the environmental conditions. Accordingly, the ultimate achievement of proteomics could be to elucidate the principles behind the molecular and functional plasticity of proteins by mapping and quantitating PTMs. Strategies that combine affinity enrichment and microfluidics for protein and peptide separation, multi-stage MS for protein quantitation and amino-acid sequencing, and bioinformatics are beginning to deliver insights into the various functional roles of PTMs in cellular processes, organelles and whole organisms.

In this article, I provide an overview of the analytical challenges that we face in proteomics studies that are aimed at the analysis of functional PTMs. I describe the concept of modification-specific proteomics, and outline recent strategies that use MS and associated technologies for the identification of PTMs in complex biological systems.

Modern proteomics

Proteomics is a rapidly expanding field that has the aim of systematically studying protein structure, function, interactions and dynamics. The pace of development is propelled by the integration of novel computational techniques, ever more advanced and sensitive analytical methods, and optimized genetics and biochemical approaches to study intricate biological systems. As proteomics technologies improve and the biological questions become more complex, increasingly advanced qualitative and quantitative methods for protein analysis are being developed[14,15].

MS is a versatile and indispensable tool in protein chemistry and proteomics[16]. The mass spectrometer determines the mass-to-charge ratio (m/z) of peptide and protein ions that are generated by electrospray ionization (ESI) or matrix-assisted laser desorption/ionization (MALDI) sources, and tandem MS (MS/MS) enables the peptides to be sequenced and PTMs to be identified and characterized. The sensitivity and accuracy of modern mass spectrometers allows the analysis of nanogram levels of individual proteins and the proteolytically derived peptides. The mass error is often less than 10 ppm (for example, 0.02 Da for a peptide of 2000 Da), but this is instrument dependent. MS-based proteomics analyses of complex protein samples usually require a starting amount of protein that is in the range of 0.05–5 mg, depending on the aims and the experimental set-up. Large amounts of protein are usually obtained from cell cultures, but it is also feasible to study protein extracts from primary tissues.

Post-translational modifications: the challenges

The investigation of PTMs presents formidable analytical challenges. More than 200 different types of PTM have been characterized[1] (FIG. 1) and new ones are regularly reported (see, for example, the DeltaMass list on The Association of Biomolecular Resource Facilities, in Further information). PTMs come in many sizes and have a wide range of physicochemical properties. They generate a large diversity of gene products because many types of PTMs are covalently attached to amino-acid residues in each protein. Mutually exclusive PTMs and heterogeneous modifications at distinct amino-acid residues lead to further complexity at the protein level (FIG. 2). As PTMs alter the molecular mass of proteins, the mapping, identification and characterization of individual modifications is often achieved by the MS analysis of an intact protein, when possible, and of the proteolytically derived peptides (BOXES 1,2).

PTMs are site-specific; they are located at specific amino-acid residues in proteins, usually in the context of a particular sequence motif (sequon). For example, *N*-linked glycosylation takes place at the Asn residue in the sequon Asn-X-Ser/Thr/Cys (where X represents any amino acid except Pro). Computational sequence analysis might therefore reveal potential sequence motifs for PTMs in proteins, including *N*-linked glycans, glycosylphosphatidylinositol (GPI) anchors and some of the most common phosphorylation sites (such as kinase-recognition motifs). However, the assignment of a PTM to a probable sequon in a protein by computational sequence analysis does not mean that the sequon is actually modified *in vivo*. Experimental methods are therefore needed to determine the actual occupancy of potentially modified sites in proteins at a given cellular stage.

Microfluidics
Nanolitre-flow chromatographic systems for protein and peptide separations prior to mass-spectrometry analysis.

Electrospray ionization
A 'soft' ionization method for mass spectrometry that generates gas-phase ions from peptide and protein solutions through the vapourization of liquid in an electric field.

Matrix-assisted laser desorption/ionization
A 'soft' ionization technique for mass spectrometry that produces gas-phase ions through the pulsed, ultraviolet laser irradiation of crystalline deposits of peptides and proteins.

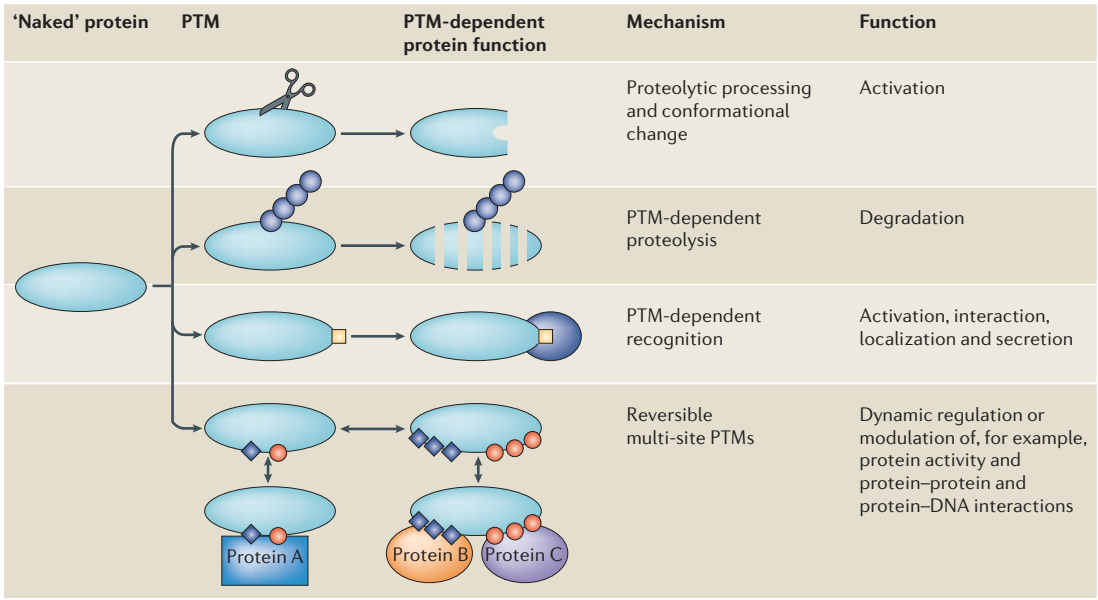


Figure 2 | **Mechanism of action of post-translational modifications.** Post-translational modifications (PTMs) can function in various ways. Proteolytic processing leads to activation by cleavage at distinct sites in a protein. In addition, PTMs might serve as the 'kiss of death' to tag proteins for destruction — for example, the polyubiquitylation-mediated degradation of proteins. Alternatively, a PTM might induce conformational changes or form a docking site to mediate molecular recognition and stabilize protein–ligand and protein–protein interactions. PTMs are also involved in the sequestration of proteins to cellular organelles and macromolecular structures. PTMs are often transient, which provides a way to rapidly and dynamically regulate protein activities, the assembly–disassembly of macromolecular complexes and the translocation of proteins between cellular compartments.

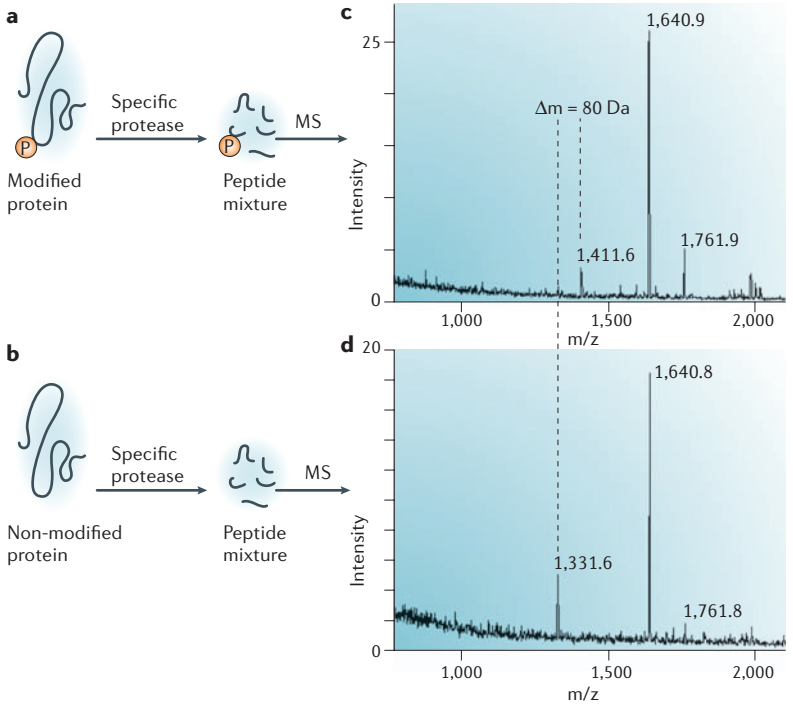
Covalently attached PTMs are usually present at sub-stoichiometric levels. This means that a PTM at a given site is often present in only a small fraction of the protein molecules of a given type. For example, the occupancy of a phosphorylation site in 5% of a protein population, while 95% of the protein population remains unmodified and inactive, might suffice to activate a signalling pathway. This creates a huge analytical challenge in proteomics because thousands of proteins are present in a biological sample. The identification of low-occupancy sites among an excess of unmodified species is being addressed by techniques that offer a wide dynamic range by combining separation and analysis technologies that provide high specificity, selectivity and sensitivity. In proteomics, these features are typically attained by combining protein-extraction methods, affinity enrichment and chromatographic and/or electrophoretic separation with peptide-mass determination and amino-acid sequencing by high-performance MS. Recently introduced quantitative MS methods might eventually determine the 'trigger level' of PTM-induced changes in protein activity and distinguish functionally important PTMs from constitutive PTMs. Reversible PTMs, such as phosphorylation and acetylation, are particularly interesting as they are dynamically regulated and are only present at distinct sites in a protein under certain conditions, depending on cellular activity and environmental factors.

Determining the presence of PTMs by western-blot analysis is a widespread and easy method for analysing protein modifications. However, western-blot analysis relies on prior knowledge of the type and position of specific modifications in proteins and on the availability of high-quality antibodies. By contrast, MS can be used to monitor known PTMs as well as to discover novel modifications. For example, in the area of chromatin structure and epigenetics, MS analysis has identified a large set of novel histone modifications, including sites for phosphorylation, ubiquitylation, methylation and acetylation[12]. As a result of such 'PTM-discovery' experiments, novel reagents and specific antibodies for the detection and quantitation of individual PTMs are being developed.

Analysing modifications using mass spectrometry

MS is ideally suited to the determination of PTMs because the covalent addition of a chemical moiety to an amino acid leads to an increase in the molecular mass of that residue, the corresponding (tryptic) peptide and the intact protein[2,3]. For example, the phosphorylation of a Tyr residue (163 Da) increases its mass to 243 Da by the addition of an HPO_3 group (80 Da). This allows the phosphorylation site to be assigned through the observation of a discrete mass increment of the intact protein or the tryptic peptide in MS or of the Tyr residue in MS/MS. MS therefore enables the site-specific assignment of PTMs at the resolution of individual amino acids in proteins. In principle, any PTM can be detected, provided that it leads to a mass increment or deficit. Prominent examples include the deamidation of Asn and Gln to Asp and Glu respectively (+1 Da), the formation of Cys–Cys disulphide bonds (–2 Da) and the addition or removal of phosphate groups (+/–80 Da) or acetyl groups (+/–42 Da) (BOXES 1,2; TABLE 1).

Box 1 | **Mapping post-translational modifications using mass spectrometry**



The detection of post-translational modifications (PTMs; represented by phosphorylation (P) in the figure) by mass spectrometry (MS) relies on the observation of mass shifts (Δm) relative to the calculated molecular mass of the 'naked' peptide or protein. A series of commonly observed PTMs are listed in TABLE 1 and an example analysis of a post-translationally modified protein is shown in the figure. Comparative MS analysis of proteolytic peptides that were derived from a purified, modified (phosphorylated) protein (see figure, part **a**) and the purified, non-modified protein (part **b**) indicates differences in the peptide mass spectra (compare parts **c** and **d**). The mass difference between the peptides at the mass-to-charge ratio (m/z) 1411.6 (part **c**) and m/z 1331.6 (part **d**) is 80 Da, and this difference corresponds to the addition of an HPO_3 moiety to the peptide. Therefore, MS analysis of relatively simple peptide mixtures — for example, those derived by the trypsin digestion of a small protein (<50 kDa) — can reveal candidate post-translationally modified peptides through accurate mass determination and a knowledge of the protein's amino-acid sequence and the protease cleavage specificity. Proteins are typically digested by trypsin and/or endoproteinase Lys C to generate peptides in the mass range 0.5–4 kDa that are suitable for MS and tandem MS (MS/MS) analysis. However, combinations of other sequence-specific or unspecific proteases can be used to obtain further amino-acid sequence information from proteins. Experimental analysis of the non-modified protein is often replaced by computational amino-acid sequence analysis and peptide mass predictions.

Although it is a useful first step, a comparative mass analysis of a peptide mixture does not provide any information on the exact location of the modified amino-acid residue in the modified peptide. MS/MS is therefore typically used as a second step to sequence candidate post-translationally modified peptides and allow the assignment of a PTM to a single residue (BOX 2).

The separation of peptides by reversed-phase capillary high-performance liquid chromatography (LC) in combination with MS/MS methods (LC–MS/MS) is useful for mapping modifications in proteins and protein complexes. MS/MS for the amino-acid sequencing of individual peptides relies on the automated, mass-specific selection and collision-induced dissociation of peptide ions inside a mass spectrometer[14]. Many post-translationally modified peptides generate distinctive modification-specific signals in MS/MS experiments, including the loss of the PTM from the intact peptide (this is known as neutral loss) or distinct ion signals that are generated by the PTM moiety. Tandem mass spectrometers can be programmed to detect and sequence only those peptides that generate such PTM-specific signals. For example, phosphoserine (pSer) is unstable and readily loses phosphoric acid in MS/MS experiments to generate a neutral loss of 98 Da from the peptide, whereas peptides containing acetyl–Lys residues generate a specific amino-acid-fragment ion signal at m/z 126.1 (BOX 2). The selective MS/MS scan modes are known as neutral-loss scans and precursor-ion scans, and they are highly useful for annotating PTMs in purified proteins[17–19]. Ion-trap-type mass analysers cannot carry out precursor-ion scans, but they offer the advantage that they can be programmed to carry out rapid multi-stage mass analysis on the timescale of peak elution in LC–MS. This allows, for example, the structural analysis of neutral loss products to confirm the presence of phosphorylation modifications in large-scale phosphoproteomics analyses[20,21] and the multi-stage MS analysis of large polypeptide fragments that are derived from intact modified proteins such as histones[22–24].

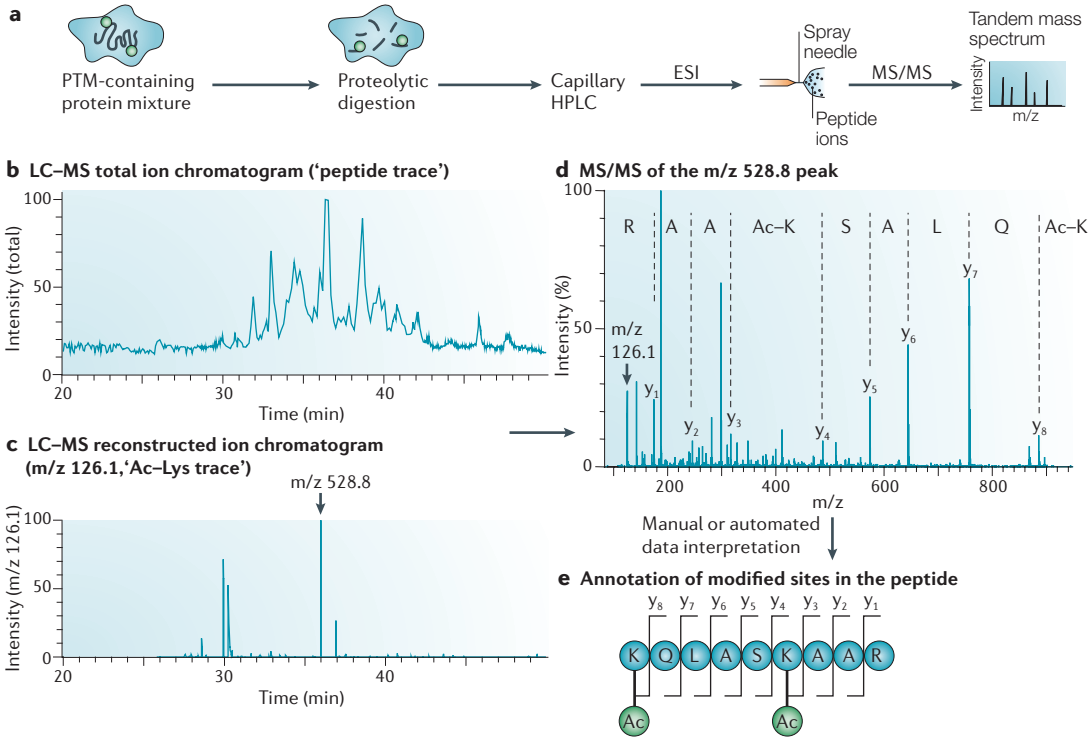
Computer algorithms for annotating PTMs using MS and MS/MS data are continuously improving and will eventually facilitate the automated annotation of PTMs[25–27]. It should be kept in mind that protein identification can be achieved by searching sequence databases using a short amino-acid sequence that has been obtained by MS/MS on a single peptide. By contrast, mapping all of the PTMs of a protein in principle requires the detection of all of the peptides of the protein (100% sequence coverage). This is not a trivial task because only a subset of the peptides that are derived from the proteolytic digestion of a protein is detected by MS analysis, unless optimization strategies are used.

Proteomics samples are exceedingly complex, often containing tens of thousands of peptides that are derived from hundreds, if not thousands, of proteins that have been extracted from a biological system. So, new and improved analytical strategies need to be implemented to allow PTMs to be determined comprehensively using proteomics approaches. Some integrated approaches for the systematic analysis of PTMs are described below.

Modification-specific proteomics

As discussed, proteomics analyses of post-translationally modified proteins face many challenges. Following the realization that post-translationally modified proteins are rarely detected or characterized in detail by 'global proteomics' approaches, more focused strategies to investigate different classes of post-translationally modified proteins have been developed[2,3,28,29]. The main aims of modification-specific proteomics are to identify proteins that carry PTMs, to map PTM sites, to quantify the changes in PTM abundance at individual sites and to characterize distinct, yet cooperative, PTMs in proteins (that is, interrelated PTMs at several sites). Biological insights can be obtained by comparing perturbed experimental systems and the relevant controls

Box 2 | **Assigning post-translational modifications using mass spectrometry**



As an example of assigning post-translational modifications (PTMs) using mass spectrometry (MS), I discuss the assignment of acetylated Lys (Ac–Lys) residues in histones. Tryptic peptides from histones were separated by nanolitre-flow high-performance liquid chromatography (HPLC) and analysed on-line by electrospray ionization (ESI)–tandem MS (MS/MS) (see figure, part **a**). MS and MS/MS spectra were automatically obtained by data-dependent acquisition. The data set was subsequently analysed to identify the proteins and determine the PTMs. Protein-sequence-database searching using the MS/MS data as the input can be used to assign post-translationally modified peptides, but complementary approaches are often required for a more thorough analysis of such data sets.

The liquid chromatography (LC)–MS data set is visualized as a total ion chromatogram ('peptide trace'), in which peptide-ion intensity is shown as a function of peptide retention time (part **b**). To detect Ac–Lys residues in this experiment, certain distinct features of this PTM were considered in the post-run data analysis (part **c**). Acetylation of a Lys residue in a peptide leads to a mass increment of +42 Da, which can be detected by MS. A mass increment of +42 Da is also evident for the modified Lys residue using MS/MS. In some cases, further information is available in the mass spectra because chemically modified amino-acid residues generate modification-specific fragment ions or they eliminate modification-specific neutral moieties (neutral loss). In the case of Ac–Lys residues, a diagnostic ion signal at the mass-to-charge ratio (m/z) 126.1 is generated from the modified peptides. Therefore, extraction of the spectra that contain the m/z 126.1 ion signal from the LC–MS data set (part **b**) will retrieve the subset of data that originated from Ac–Lys peptides ('Ac–Lys trace') (part **c**). The MS/MS spectrum of a candidate Ac–Lys peptide at m/z 528.8 that elutes at 36 min (part **c**) indeed contains the m/z 126.1 signal (part **d**). The ladder-like y-ion series of peptide-fragment ions facilitated the assignment of the amino-acid sequence and the modified sites (part **d**). This MS/MS spectrum allowed the assignment of two Ac–Lys residues in the histone protein (part **e**).

to produce temporal or functional abundance profiles of post-translationally modified proteins.

A key to improving proteomics approaches in the area of PTMs is to purify the organelles or protein species of interest prior to protein and PTM characterization by MS/MS. The most successful strategies integrate various modification-specific enrichment steps in the protocols or use combinations of protein-digestion and peptide-separation methods. Some methods use chemical reactions to capture PTMs, whereas other methods rely on genetic, biochemical or chromatographic methods to specifically recover post-translationally modified proteins and peptides. The various stages of modification-specific proteomics are described in BOX 3.

Using affinity enrichment. The enrichment of modified proteins is often achieved by immunoprecipitation using protein-specific antibodies. Individual proteins can be selectively purified and digested and then analysed by LC–MS/MS. When using antibodies, it is important to consider that PTMs might mask or alter the protein epitope that is normally recognized by the antibody, which could lead to a reduced recovery of the post-translationally modified protein in immunoprecipitation experiments. To analyse the phosphorylation of plasma-membrane receptors, it is advisable to use antibodies that are targeted to the extracellular part of the protein. This approach was recently used in a systematic analysis of the epidermal growth factor

Table 1 | **A selection of post-translational modifications, their relevant mass values and biological roles**

Post-translational modification* (modified residues)	Change in mass (Δm, Da)	Example biological functions and comments
Phosphorylation		
(Ser, Thr, Tyr)	80	Signal transduction, regulation of enzyme activity, involved in protein–protein and protein–ligand interactions; phosphorylation and sulphation have the same Δm (80 Da), but can be distinguished by positive- or negative-ion mass-spectrometry analysis
Glycosylation		
N-linked (Asn) *O*-linked (Ser, Thr) GPI anchor	> 800 203, > 800‡ > 1,000	Protein stability, solubility, secretion signal, regulator of interactions, extracellular recognition and interactions; modification by a GPI anchor is coupled to protein processing
Acylation		
Palmitoylation Farnesylation Myristoylation	238 204, 206§ 210	Protein localization and activity, involved in protein–protein and protein–membrane interactions
Sulphation		
(Tyr)	80	Signalling and protein localization, involved in protein–protein interactions; phosphorylation and sulphation have the same Δm (80 Da), but can be distinguished by positive- or negative-ion mass-spectrometry analysis
Ubiquitylation		
(Lys)	> 1,000	Protein degradation signal, involved in protein–protein interactions; observed as a +114-Da mass tag (Gly–Gly) after the tryptic digestion of a modified protein
Methylation		
(Lys mono-, di- and trimethylation, Arg mono- and dimethylation)	14, 28, 42	Regulates protein activity, protein–protein and protein–nucleic-acid interactions, chromatin dynamics and gene activity (histone modification); note that trimethylation and acetylation have similar mass increments
Acetylation		
(N-terminal residue, Lys)	42	Protein stability and activity, regulates protein–protein and protein–ligand interactions; note that trimethylation and acetylation have similar mass increments
Disulphide-bond formation		
(Cys)	–2	Stabilizes protein structure and activity, involved in redox processes
Oxidation		
(Met) (Trp)	16 4, 16, 32‖	Might regulate protein activity; often a chemical artefact
Deamidation		
(Asn, Gln)	1	Associated with ageing, might regulate protein activity and interactions; often a chemical artefact
Hydroxylation		
(Pro)	16	Structural stability (collagens)

*A more comprehensive list of post-translational modifications can be found on The Association of Biomolecular Resource Facilities and UNIMOD web sites (see Further information). ‡*O*-linked glycosylation can be a single sugar (203) or a chain of sugars (>800). §The Δm for farnesylation depends on the degree of saturation of the acyl group. ‖Different oxidized products can be generated from Trp. GPI, glycosylphosphatidylinositol.

(EGF) receptor under various cellular conditions that revealed stimulation-dependent multi-site differential phosphorylation[30].

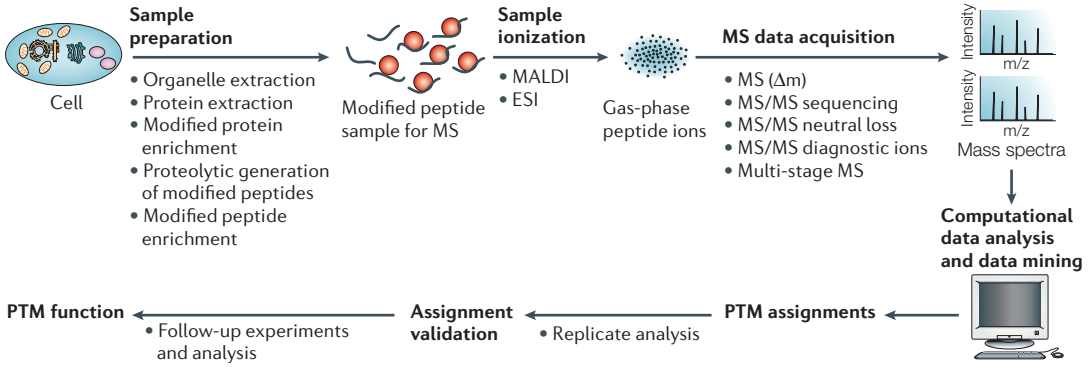
Modification-specific antibodies are highly useful for enriching the corresponding modified proteins. Commercially available phosphotyrosine (pTyr)-specific antibodies are widely used for the investigation of cell-signalling processes. The differential analysis of stimulated and unstimulated cells by pTyr-specific immunoprecipitation and protein identification by MS has revealed a range of signalling factors that become activated following growth-factor treatment[31–33]. Also, the recent successful application of pTyr-specific antibodies to the enrichment of phosphorylated peptides in

Box 3 | **The overall modification-specific proteomics strategy**

Most functionally interesting post-translational modifications (PTMs) are present at only low levels in cells and tissues. Modification-specific analytical strategies therefore aim to achieve high sensitivity and specificity and a wide dynamic range for the analysis of complex biological samples, so that low-abundance and low-stoichiometry PTMs are detected[2]. Analysis is often targeted towards a single type of PTM — for example, towards phosphotyrosine-, phosphoserine- and phosphothreonine-containing proteins or *N*-glycosylated proteins. In addition, PTM-targeted methods can be combined with organelle-targeted methods[15] to obtain biological insights about regulatory events in subcellular compartments.

The analytical strategies involved often combine PTM-specific affinity-enrichment techniques with electrophoretic and microfluidic separations, advanced mass spectrometry (MS) and bioinformatics (see figure). The important analytical stages include: organelle or protein extraction and/or post-translationally modified protein enrichment, followed by the proteolytic generation of peptides; post-translationally modified peptide enrichment using affinity or chemical methods; optimal peptide sample preparation for MS analysis; tandem MS (MS/MS) analysis and the sequencing of post-translationally modified peptides using advanced features, such as data-dependent acquisition and multi-stage MS; data analysis and data mining using advanced bioinformatics tools; the validation of results by replicate analysis; and, last, follow-up biological experiments and bibliographic analyses. Such PTM-specific methods can be combined with protein-quantitation techniques, including stable-isotope labelling and peptide-intensity profiling (BOX 4).

It should be noted that sample preparation has a crucial role in the proteomics analysis of PTMs. A large fraction of PTMs might be missed when cellular conditions are not strictly controlled or when the protein-extraction method introduces artefacts in the form of endogenous active enzymes (for example, phosphatases or glycosidases) that remove or modify PTMs. Further problems might arise during protein and peptide separation and MS analysis — for example, the loss of post-translationally modified peptides owing to their physicochemical properties (such as hydrophobicity or polarity) or the elimination of PTMs owing to pH changes or a suboptimal set-up of the MS experiment. Δm, mass shift; ESI, electrospray ionization; MALDI, matrix-assisted laser desorption/ionization; m/z, mass-to-charge ratio.



cancer proteomics has been reported[34]. Kinase-substrate-specific pSer- and phosphothreonine (pThr)-specific antibodies are also available and might be applied to the study of phosphoproteomes[35,36]. Unfortunately, however, only a few efficient modification-specific antibodies are currently available for the targeted immunoprecipitation of modified proteins. A larger selection of modification-specific antibodies would benefit proteomics tremendously.

Phosphopeptides can be enriched by immobilized metal-affinity chromatography (IMAC), which takes advantage of the affinity of the phosphate moiety for Fe(III) or Ga(III)[37,38]. Despite debates over its efficiency, the IMAC method has been successfully applied in large-scale phosphoproteomics studies that have identified hundreds of phosphorylated residues in proteins from various species[21,39–41]. Under optimized conditions, IMAC is also useful for the enrichment of intact phosphoproteins in proteomics experiments[42]. Some pre-fractionation of the proteins or peptides prior to IMAC improves the selectivity of the method. Recently, TiO_2 columns were introduced for phosphopeptide enrichment prior to MS analysis[43,44], and they seem to be promising tools in functional phosphoproteomics studies.

TiO_2 columns
Titanium dioxide interacts with phosphate groups by a chelation mechanism. These columns are therefore useful for the selective and specific enrichment of phosphopeptides.

Glycoproteins are abundant in cells, in the plasma membrane and in body fluids. The glycosylation of proteins that are present at the cell surface provides a means for cell–cell recognition and communication. Glycan structures distinguish healthy and diseased cells and tissues, and are therefore of interest for diagnostic and therapeutic biomedicine[45]. Proteomics analysis of glycoproteins by MS is only in its infancy, but several promising methods have been introduced[46,47]. Lectins, which are often considered to be carbohydrate antibodies, have distinct specificities towards various classes of glycan and are generally used to enrich glycoproteins and glycopeptides[48]. Combining various types of lectin can improve the selectivity and specificity of proteomics experiments that are aimed at investigating *N*-glycosylated and *O*-glycosylated proteins and peptides[49]. In a large-scale experiment to delineate the glycoproteins in *Caenorhabditis elegans*, a lectin was first used to enrich the glycoproteins in a protein extract. This was followed by the trypsin digestion of the glycoproteins and a subsequent lectin-mediated enrichment of the resulting glycopeptides[50]. Following endoglycosidase (specifically, peptide–*N*-glycosidase F or PNGase F) treatment in ^{18}O-water to generate an

N-glycosylation-specific mass tag on the modified peptides, the glycosylation sites were identified by MS[50,51]. Combining the lectin-mediated affinity purification of glycoproteins with hydrophilic-interaction liquid chromatography (HILIC) for the recovery of glycopeptides prior to MS is another useful method for the analysis of complex glycoprotein samples such as plasma or serum[52].

Glycans that are enzymatically or chemically released from glycopeptides can be analysed and characterized by MS[47]. SDS–PAGE-based or two-dimensional (2D)-gel-based strategies for the detailed MS analysis of carbohydrate structures at individual glycosylation sites have been reported[46,53,54], but the achievable sensitivity remains an obstacle for large-scale glycoproteomics studies.

Using chemistry. Affinity-based methods can be complemented by various chemical methods for the specific capture, immobilization and alteration of post-translationally modified peptides to improve the overall dynamic range, sensitivity and specificity of the analysis. The solid-phase capture of PTMs provides a way to carry out chemical and biochemical reactions followed by the release and analysis of the altered post-translationally modified peptides by MS. *O*-phosphorylated or *O*-glycosylated Ser and Thr residues are typically converted to more tractable species by using β-elimination and Michael addition reactions[55]. Phosphotyrosine peptides can be captured on solid supports using phosphoramidate chemistry[56], which involves coupling them to polymer (specifically to dendrimer) resins[57]. Fluorous affinity tags were recently introduced as an elegant way to enrich post-translationally modified peptides, including pSer peptides, by converting them into perfluorinated species that could be efficiently recovered by solid-phase extraction and analysed by MS[58]. It should be kept in mind that methods that rely on β-elimination, Michael addition or other types of chemical conversion of modified amino-acid residues are often hampered by technical difficulties and the generation of by-products. Also, the adaptation of these methods for large-scale modification-specific proteomics requires further developments to improve their efficiency, sensitivity and versatility.

'Tagging-via-substrate' techniques have also entered the field of modification-specific proteomics. Protein modification by *O*-linked *N*-acetylglucosamine and farnesylation was investigated by combining the metabolic incorporation of PTM-precursor analogues and chemo-selective activation for the solid-phase capture of the post-translationally modified species after their extraction from the cell[59,60]. Cells were grown in a medium that contained an azido analogue of a metabolic precursor for the PTM in question. The analogue was metabolically incorporated into proteins and, following the specific activation of the PTM analogue, it reacted with biotin, which enabled the selective recovery of the modified proteins from the cell lysates by streptavidin-affinity enrichment and their subsequent analysis by MS/MS.

The high selectivity of chromatographic stationary phases is explored in the combined fractional diagonal chromatography (COFRADIC) method, which allows the identification of modified peptides by duplicate (diagonal) LC analysis that is intersected by a chemical or enzymatic reaction[61]. The method is based on the idea that the chemical modification of peptides affects their retention time during LC. In a first step, peptides are separated by LC and collected in discrete fractions. Individual peptide fractions are then treated with a PTM-specific reagent (which induces, for example, *S*- or *N*-alkylation) that can change the physicochemical properties of only the modified peptides to increase or decrease their hydrophobicity. A second LC separation then shifts the chemically or enzymatically converted peptides to a different retention time — that is, away from unmodified peptides — which facilitates their recovery for further analysis by MS. COFRADIC facilitates the systematic analysis of the N termini of proteins and thereby provides a unique way to study protein processing on a proteomics scale, as has recently been shown in a study of caspase substrates[62].

The selectivity of strong cation-exchange chromatography (SCX) is useful for the isolation of phosphopeptides. The presence of the phosphate group reduces the isoelectric point of phosphopeptides relative to the unmodified species and results in an enrichment of phosphopeptides in the early eluting SCX fractions[20]. Similarly, some *N*-acetylated peptides are also found in SCX fractions that elute early.

Large modifications: ubiquitin and related modifiers. Proteins are not only modified by small, chemical moieties, but are also post-translationally modified through their conjugation to other proteins. Ubiquitin is a small protein of 76 amino acids that is often covalently attached to specific Lys residues in substrate proteins by ubiquitin-ligase enzymes[63]. Polyubiquitylation can target proteins for proteasome-mediated degradation and thereby has an important function in the regulation of protein abundance and turnover in cells. Ubiquitin-like modifiers — such as SUMO (small ubiquitin-related modifier), NEDD8 (neuronal-precursor-cell-expressed developmentally downregulated protein-8) and ISG15 (interferon-stimulated gene-15) — are polypeptides that are also conjugated to proteins and are involved in the regulation of a range of cellular processes and pathways[63].

Proteomics analyses of ubiquitylation and modification by ubiquitin-like modifiers typically include the genetic tagging of ubiquitin or ubiquitin-like modifiers, followed by affinity purification and protein identification by MS[29]. Using this approach, hundreds of ubiquitin and SUMO targets have been identified in various organisms, which has revealed a range of new substrates and potential novel functions for these PTMs. A proteomics screen for ISG15 substrates in mouse and human cells was carried out using immunoprecipitation and the MS-based identification of ISG15-conjugated proteins[64].

β-elimination reaction
The alkaline-induced elimination of phosphoric acid from phosphoserine and phosphothreonine residues or of *O*-linked sugars from *O*-glycosylated residues in peptides.

Michael addition reaction
The addition of a functional chemical moiety to a residue that was post-translationally modified but that has undergone a β-elimination reaction. Affinity tags or mass tags can be added to allow the specific recovery or detection of post-translationally modified peptides.

Phosphoramidate chemistry
A method for the selective capture of converted post-translationally modified peptides, with the aim of identifying the modified protein and the site of the modification.

Chromatographic stationary phases
The column resins that are used for the chromatographic separation of molecules — for example, a reversed-phase resin or a strong-cation-exchange resin.

Small modifications: redox processes. Oxidative processes are important for cellular survival and they are also involved in a series of pathological conditions, including inflammation. Reactive oxygen species and reactive nitrogen species are now recognized as important regulators of protein function and turnover, and a number of proteomics strategies for the determination of oxidative events have been reported[65]. For example, the nitration of Tyr residues generates 3-nitrotyrosine, which functions as a marker for oxidative stress and ageing. The 'nitroproteome' of organs, cells and organelles can be investigated by gel electrophoresis, western blotting using a nitrotyrosine-specific antibody and MS[66,67]. A proteomics analysis of rat retinas revealed that nitration might have a role in the light adaptation of vision[68]. Using a different proteomics strategy, MS was used to identify nitrated proteins and to locate nitrotyrosine residues in the ageing rat muscle[69].

Analysing the modifications of organelles. The isolation of organelles, subcellular fractions or plasma-membrane preparations is a useful first stage of sample preparation in modification-specific proteomics (BOX 3). This greatly reduces sample complexity and increases the probability of identifying biologically significant PTMs. The LC–MS/MS analysis of peptides derived from cellular membrane fractions and protein complexes has revealed modified proteins, including oxidized species[70,71]. Targeted large-scale phosphoproteomics analyses of HeLa cell nuclear preparations[20] and mouse synapses[42] have identified thousands of protein phosphorylation sites. Proteomics analyses of plasma-membrane proteins and their PTMs are at centre stage in many areas of molecular cell biology, from studies of invasive pathogenic microorganisms to searches for stem-cell differentiation markers. The modification-specific 'shave-and-conquer' strategy takes advantage of specific proteases or lipases and detergent-based separation methods for the selective release and identification of distinct classes of post-translationally modified proteins, including phosphoproteins and GPI-anchored proteins, from plasma-membrane preparations[40,72].

Functional phosphoproteomics and cell signalling. Cell-signalling events are transient and propagate from receptors at the plasma membrane to receiver and effector molecules in the cytoplasm, nucleus and other organelles. Functional phosphoproteomics aims to identify and characterize signalling molecules and their phosphorylated domains to elucidate the mechanistic details of signal propagation in cellular networks. Therefore, qualitative and quantitative methods are used to define the post-translationally modified proteins.

Protein quantitation by MS is widely used in protein-expression studies, and such methods were recently used for the quantitative analysis of PTMs in functional proteomics analyses of biological systems. Two main technologies have emerged — peptide-intensity profiling (PIP) and stable-isotope labelling (SIL) of peptides and proteins (BOX 4). PIP relies on the accurate comparison of the results of many LC–MS experiments. Usually, the results of a series of LC–MS experiments are compared and aligned on the basis of peptide-retention time and accurate peptide-mass determination. Computational normalization and comparative data analysis facilitate the determination of quantitative changes in peptide intensities across many experiments (BOX 4). The PIP analysis of protein phosphorylation using a comparative LC–MS analysis of phosphopeptides is possible when appropriate correction factors are considered[73]. It remains to be seen how applicable such methods are to large-scale phosphoproteomics studies.

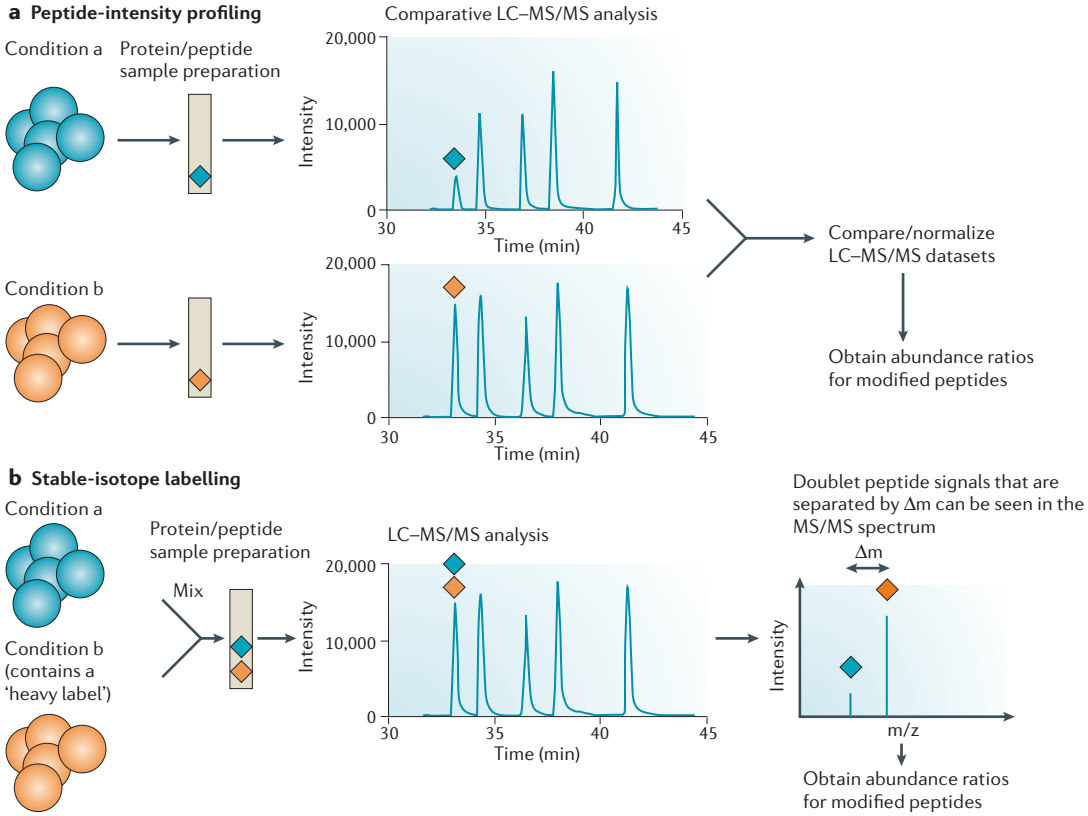
SIL relies on the introduction of stable-isotope labelled amino-acid precursors or chemical reagents into proteins or peptides[14,16,74,75]. The metabolic or chemical incorporation of ^{15}N, ^{13}C, ^{18}O or ^{2}H into peptides or proteins allows the labelled peptides to be distinguished from their unlabelled counterparts by a mass increment that relates to the number of heavy atoms that were introduced (BOX 4). Using stable-isotope-labelled internal peptide standards, it is also feasible to achieve absolute quantitation of modified peptides and proteins[76].

In the context of PTMs, it is possible to use modification-specific chemistries and SIL alkylating reagents to label all of the candidate post-translationally modified peptides specifically at their PTM sites. Another possibility is the comprehensive labelling of peptides using a SIL alkylating reagent or by incorporating ^{18}O during proteolysis followed by PTM-specific affinity enrichment. The use of stable-isotope-tagged amine-reactive reagents (for example, iTRAQ; Applied Biosystems) allows up to four experimental states to be monitored in one analysis[77]. This multiplexing technology was used in combination with pTyr immunoprecipitation, IMAC and MS/MS in a time-resolved study of the EGF signalling cascade. By labelling tryptic peptides from four different EGF-receptor stimulation time points with four isoforms of the iTRAQ reagent to enable downstream quantification, this approach allowed the identification of signalling modules in a series of proteins[41].

The metabolic SIL of proteins by ^{15}N incorporation in *Saccharomyces cerevisiae* allowed the site-specific identification and quantitation of protein phosphorylation by the kinase Ste20[74]. Since then, SIL using metabolic ^{15}N incorporation has been refined and used in other model organisms, including *C. elegans* and *Drosophila melanogaster*[78], to enable comprehensive quantitative proteomics approaches. A variation of metabolic SIL, which is known as stable-isotope labelling by amino acids in cell culture (SILAC), takes advantage of stable-isotope-labelled amino acids — for example, $^{13}C_6$-Arg and $^{13}C_6$-Lys — to allow quantitative proteomics studies to be carried out using MS and a range of eukaryotic cell lines[21,79–82]. The combination of SILAC, immunoprecipitation and MS was successfully used both to investigate PTM-dependent temporal and functional aspects of cell signalling during the differentiation of human cells[33,83] and for the detailed investigation of cell-signalling cassettes[84]. Using SILAC, the affinity enrichment of phosphopeptides and multi-stage MS, a large-scale

Cell-signalling cassettes
Multiprotein modules that are involved in signal-transduction processes and that rapidly propagate stimuli from cell-surface receptors to the cytosol or the nucleus.

Box 4 | **Quantitative proteomics analyses of post-translational modifications**



The accurate quantitation of protein-expression levels and post-translational modifications (PTMs) is a prerequisite for obtaining molecular and functional insights into dynamic cellular processes, including the spatial and temporal organization of cellular networks and metabolic processes. Mass spectrometry (MS)-based quantitative proteomics relies on the measurement of the masses and intensities of peptides that are derived from cellular proteins. The electrospray ionization MS response of a peptide (the peptide-ion intensity) correlates with the concentration of that peptide in the sample. It is therefore possible to obtain relative quantitative data for the abundance variations of individual peptides by using a comparative analysis (peptide-intensity profiling (PIP); see figure, part **a**). Liquid-chromatography–tandem mass spectrometry (LC–MS/MS) is well suited to comparative quantitative analyses of peptide mixtures because the peptide retention time during the LC separation and the peptide mass that is determined by MS are unique identifiers that can be used to align data sets. The accuracy of the comparative analysis therefore relies on the chromatographic system and the mass spectrometer, specifically on factors such as the reproducibility of the LC gradient, the robustness of the the ion source and the mass calibration of the MS instrument.

Stable-isotope labelling (SIL) is useful for pairwise comparisons of samples (see figure, part **b**). Proteins in one sample contain the natural element ('light label'), whereas the other protein sample contains a stable-isotope-labelled reagent or amino acid ('heavy label'). For example, the introduction of $^{13}C_6$-Lys into proteins leads to a mass increment for all of the labelled Lys-containing peptides relative to the same peptides that were obtained from the unlabelled light protein sample. Therefore, the two samples can be mixed and processed in one experiment. Tryptic digestion and LC–MS/MS analysis gives a readout of Lys-containing peptide pairs (light/heavy ratio), and their intensity ratios reflect the relative abundance of the peptides in the two systems. The combination of PIP or SIL with PTM-specific affinity enrichment allows the quantitation of post-translationally modified peptides. The multiplex analysis of PTMs can be carried out using 'triple' labels (for example, by metabolically labelling the proteomes of three cell populations with different stable isotopic forms of arginine[33]) and iTRAQ™ chemistry (for example, by labelling the tryptic peptides isolated from cells under four different conditions with four isoforms of the iTRAQ reagent to enable downstream quantification[41]). Δm, mass shift; m/z, mass-to-charge ratio.

quantitative phosphoproteomics study of the pheromone response in *S. cerevisiae* revealed more than 130 functionally relevant phosphorylation sites among a total of 700 phosphorylation sites[21]. These functional phosphoproteomics studies show that it is possible to carry out an in-depth analysis of PTM-dependent cell-signalling proteins, from plasma-membrane receptors to nuclear effector proteins.

Trends and new technologies

A range of genetic and biochemical approaches for the large-scale analysis of PTMs are evolving. For example, a yeast two-hybrid screen for the analysis of modification-specific binary protein interactions has been developed[85], and proteome arrays are being developed to assay protein-kinase activity[86]. Also, ChIP-on-chip methods for the genome-wide analysis of DNA-binding proteins and

their PTMs will eventually be feasible and affordable. At the same time, 2D-gel electrophoresis remains a dependable method for the separation of heterogeneously modified intact proteins, and new technologies for the visualization and quantitation of post-translationally modified proteins on 2D gels are continuously being developed[87,88].

MS will continue to have an important role in modification-specific proteomics approaches. Novel types of instrument are being regularly launched and new experimental concepts are continuously being introduced. High-resolution instruments that provide an improved mass accuracy are entering the field[89,90]. At the same time, new generations of tandem mass spectrometers are offering an improved sequencing speed and advanced multi-stage MS features for the detailed analysis of post-translationally modified peptides. The combination of high-performance MS with the systematic immunoprecipitation of sets of target proteins might eventually provide a suitable platform for large-scale functional analyses of PTMs. Large-scale efforts are underway to facilitate such initiatives by producing a library of antibodies to all human proteins[91].

Several new methods for the MS fragmentation of peptides and intact proteins have been introduced. Electron capture dissociation (ECD) facilitates the efficient amino-acid sequencing of post-translationally modified peptides and small proteins while keeping the PTMs intact[92,93]. More recently, the seemingly more efficient technique of electron transfer dissociation (ETD) was applied to the analysis of phosphopeptides and histones[22,94].

The ability to carry out multi-stage MS analysis and the availability of new and improved ion-dissociation methods, such as ECD and ETD, will produce new analytical strategies and applications that can sequence large peptides (~3–10 kDa), which will allow more comprehensive studies of multi-site PTMs in proteins.

The feasibility of carrying out a complete top-down analysis of intact modified proteins inside a mass spectrometer has already been shown, and efforts to integrate this technology into proteomics workflows are underway[95,96]. Top-down proteomics is based on the idea of introducing intact proteins into the mass spectrometer and then applying efficient gas-phase ion isolation and dissociation methods to fragment the proteins — that is, to carry out 'gas-phase proteolysis'. Proteins are dissected piece by piece in a top-down fashion by using multi-stage MS for amino-acid sequencing and the mapping of modifications. This approach relies on high-performance instrumentation that has the capacity, dynamic range, resolving power and mass accuracy to analyse highly complex populations of proteins, their fragments and various PTMs.

Integrative computational approaches

As experimental proteomics techniques provide increasingly accurate qualitative and quantitative data on PTM sites in proteins, computational techniques are being developed to process, integrate and apply this new knowledge. Proteomics standards initiatives promise to deliver guidelines and schemes for the storage and exchange of proteomics data sets[97,98]. Database annotations are steadily improving, and new databases are emerging that integrate a wealth of information on proteins and their modifications[99]. Computational sequence-analysis methods for the prediction of PTM sites in proteins have improved over recent years[100], and the current influx of novel PTM data will no doubt lead to much improved predictions in the future. In turn, sequence-based prediction methods can be used to direct the rational analysis of PTMs in distinct proteins through hypothesis-driven MS experiments to map, for example, phosphorylation sites in proteins[19,101,102]. These methods rely on prior knowledge of a protein sequence and on the prediction of putative modified sites and the corresponding peptide masses. The data-acquisition software of the mass spectrometer can then be programmed to monitor and sequence all of the predefined candidate modified peptides, which improves the analytical specificity significantly.

The functional proteomics analysis of PTMs also provides a framework for discovering novel PTM sequons and features. Amino-acid sequence analysis of large-scale phosphoproteomics data sets can reveal new sequence motifs that are phosphorylated and identify novel conserved features in protein families[103]. Using network analysis and self-organizing mapping approaches, temporal protein-phosphorylation data can reveal novel signalling domains and features of signalling cascades[33,41,104]. Integrating such studies with efforts to correlate gene expression and protein–protein interaction data[105,106] will soon make the modelling of dynamic signalling pathways and molecular networks in cells possible. Eventually, accurate and time-resolved large-scale data on PTMs will provide a foundation for computational simulations of cell-signalling processes to facilitate virtual tests of perturbations to molecular networks.

Conclusions

Over the past decade, researchers have witnessed the amazing evolution of computational and experimental technologies in molecular cell biology. At the same time, it has become evident that proteins are much more complex, diverse and dynamic than was originally anticipated. Clearly, proteomics technologies have to develop further to tackle the issues surrounding the intricate control and regulation of protein functions, localizations and interactions in cells and tissues.

Using affinity-enrichment techniques and quantitative protein and peptide analysis by MS, functional modification-specific proteomics is beginning to reveal the molecular features of complex cellular networks and will soon contribute more details to our understanding of molecular recognition and cell-signalling processes. The integration of computational tools into modification-specific proteomics studies is a prerequisite for the interpretation of large-scale data sets into meaningful biological information. With current technologies, researchers are revealing many details about PTM-mediated processes in cells, but the field still has many challenges to overcome to decipher the various dialects of the protein language.

1. Krishna, R. G. & Wold, F. Post-translational modification of proteins. *Adv. Enzymol. Relat. Areas Mol. Biol.* **67**, 265–298 (1993).
2. Jensen, O. N. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* **8**, 33–41 (2004).
3. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nature Biotechnol.* **21**, 255–261 (2003).
4. Eichler, J. & Adams, M. W. Posttranslational protein modification in *Archaea*. *Microbiol. Mol. Biol. Rev.* **69**, 393–425 (2005).
5. Yang, X. J. Multisite protein modification and intramolecular signaling. *Oncogene* **24**, 1653–1662 (2005).
An excellent review of multi-site PTMs and their biological roles.
6. Cohen, P. The regulation of protein function by multisite phosphorylation — a 25 year update. *Trends Biochem. Sci.* **25**, 596–601 (2000).
7. Gunawardena, J. Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proc. Natl Acad. Sci. USA* **102**, 14617–14622 (2005).
8. Cosgrove, M. S., Boeke, J. D. & Wolberger, C. Regulated nucleosome mobility and the histone code. *Nature Struct. Mol. Biol.* **11**, 1037–1043 (2004).
9. Fischle, W., Wang, Y. & Allis, C. D. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475–479 (2003).
10. Pokholok, D. K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517–527 (2005).
11. Fischle, W. *et al.* Regulation of HP1–chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438**, 1116–1122 (2005).
12. Freitas, M. A., Sklenar, A. R. & Parthun, M. R. Application of mass spectrometry to the identification and quantification of histone post-translational modifications. *J. Cell. Biochem.* **92**, 691–700 (2004).
13. Insinga, A., Minucci, S. & Pelicci, P. G. Mechanisms of selective anticancer action of histone deacetylase inhibitors. *Cell Cycle* **4**, 741–743 (2005).
14. Steen, H. & Mann, M. The abc's (and xyz's) of peptide sequencing. *Nature Rev. Mol. Cell Biol.* **5**, 699–711 (2004).
A good introduction to the concepts and methods that are involved in peptide sequencing by MS/MS.
15. Yates, J. R. 3rd, Gilchrist, A., Howell, K. E. & Bergeron, J. J. Proteomics of organelles and large cellular structures. *Nature Rev. Mol. Cell Biol.* **6**, 702–714 (2005).
16. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
17. Loughrey Chen, S. *et al.* Mass spectrometry-based methods for phosphorylation site mapping of hyperphosphorylated proteins applied to Net1, a regulator of exit from mitosis in yeast. *Mol. Cell. Proteomics* **1**, 186–196 (2002).
18. Neubauer, G. & Mann, M. Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: potentials and limitations. *Anal. Chem.* **71**, 235–242 (1999).
19. Unwin, R. D. *et al.* Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. *Mol. Cell. Proteomics* **4**, 1134–1144 (2005).
20. Beausoleil, S. A. *et al.* Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl Acad. Sci. USA* **101**, 12130–12135 (2004).
21. Gruhler, A. *et al.* Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics* **4**, 310–327 (2005).
A large-scale quantitative phosphoproteomics study using affinity enrichment and multi-stage MS reveals some of the key signalling modules that are involved in the yeast pheromone response.
22. Coon, J. J. *et al.* Protein identification using sequential ion–ion reactions and tandem mass spectrometry. *Proc. Natl Acad. Sci. USA* **102**, 9463–9468 (2005).
23. Sze, S. K., Ge, Y., Oh, H. & McLafferty, F. W. Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl Acad. Sci. USA* **99**, 1774–1779 (2002).
24. Wu, S. L., Kim, J., Hancock, W. S. & Karger, B. Extended range proteomic analysis (ERPA): a new and sensitive LC–MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of β-casein and epidermal growth factor receptor (EGFR). *J. Proteome Res.* **4**, 1155–1170 (2005).
25. Creasy, D. M. & Cottrell, J. S. Error tolerant searching of uninterpreted tandem mass spectrometry data. *Proteomics* **2**, 1426–1434 (2002).
26. Matthiesen, R., Trelle, M. B., Højrup, P., Bunkenborg, J. & Jensen, O. N. VEMS 3.0: algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. *J. Proteome Res.* **4**, 2327–2337 (2005).
27. Sadygov, R. G., Cociorva, D. & Yates, J. R. 3rd. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nature Methods* **1**, 195–202 (2004).
28. Cantin, G. T. & Yates, J. R. 3rd. Strategies for shotgun identification of post-translational modifications by mass spectrometry. *J. Chromatogr. A* **1053**, 7–14 (2004).
29. Kirkpatrick, D. S., Denison, C. & Gygi, S. P. Weighing in on ubiquitin: the expanding role of mass-spectrometry-based proteomics. *Nature Cell Biol.* **7**, 750–757 (2005).
Excellent review on proteomics approaches for the study of ubiquitylation and protein modification by ubiquitin-like modifiers.
30. Boeri Erba, E. *et al.* Systematic analysis of the epidermal growth factor receptor by mass spectrometry reveals stimulation-dependent multisite phosphorylation. *Mol. Cell. Proteomics* **4**, 1107–1121 (2005).
31. Pandey, A. *et al.* Analysis of receptor signaling pathways by mass spectrometry: identification of Vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc. Natl Acad. Sci. USA* **97**, 179–184 (2000).
32. Salomon, A. R. *et al.* Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc. Natl Acad. Sci. USA* **100**, 443–448 (2003).
33. Blagoev, B., Ong, S. E., Kratchmarova, I. & Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nature Biotechnol.* **22**, 1139–1145 (2004).
A temporal analysis that used triplex stable-isotope labelling and MS to study EGF-induced phosphotyrosine signalling events.
34. Rush, J. *et al.* Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nature Biotechnol.* **23**, 94–101 (2005).
35. Gronborg, M. *et al.* A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol. Cell. Proteomics* **1**, 517–527 (2002).
36. Kane, S. *et al.* A method to identify serine kinase substrates. Akt phosphorylates a novel adipocyte protein with a Rab GTPase-activating protein (GAP) domain. *J. Biol. Chem.* **277**, 22115–22118 (2002).
37. Posewitz, M. C. & Tempst, P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* **71**, 2883–2892 (1999).
38. Stensballe, A., Andersen, S. & Jensen, O. N. Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics* **1**, 207–222 (2001).
39. Ficarro, S. B. *et al.* Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nature Biotechnol.* **20**, 301–305 (2002).
40. Nuhse, T. S., Stensballe, A., Jensen, O. N. & Peck, S. C. Large-scale analysis of *in vivo* phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol. Cell. Proteomics* **2**, 1234–1243 (2003).
41. Zhang, Y. *et al.* Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell. Proteomics* **4**, 1240–1250 (2005).
42. Collins, M. O. *et al.* Proteomic analysis of *in vivo* phosphorylated synaptic proteins. *J. Biol. Chem.* **280**, 5972–5982 (2005).
43. Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B. & Heck, A. J. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC–ESI–MS/MS and titanium oxide precolumns. *Anal. Chem.* **76**, 3935–3943 (2004).
44. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P. & Jorgensen, T. J. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* **4**, 873–886 (2005).
45. Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation — potential for therapeutics and diagnostics. *Nature Rev. Drug Discov.* **4**, 477–488 (2005).
46. Wilson, N. L., Schulz, B. L., Karlsson, N. G. & Packer, N. H. Sequential analysis of *N*- and *O*-linked glycosylation of 2D-PAGE separated glycoproteins. *J. Proteome Res.* **1**, 521–529 (2002).
47. Harvey, D. J. Proteomic analysis of glycosylation: structural determination of *N*- and *O*-linked glycans by mass spectrometry. *Expert Rev. Proteomics* **2**, 87–101 (2005).
A comprehensive overview of MS-based methods for glycoprotein and glycan analysis using proteomics approaches.
48. Gabius, H. J., Andre, S., Kaltner, H. & Siebert, H. C. The sugar code: functional lectinomics. *Biochim. Biophys. Acta* **1572**, 165–177 (2002).
49. Yang, Z. & Hancock, W. S. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J. Chromatogr. A* **1053**, 79–88 (2004).
50. Kaji, H. *et al.* Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify *N*-linked glycoproteins. *Nature Biotechnol.* **21**, 667–672 (2003).
51. Küster, B. & Mann, M. ^{18}O-labeling of *N*-glycosylation sites to improve the identification of gel-separated glycoproteins using peptide mass mapping and database searching. *Anal. Chem.* **71**, 1431–1440 (1999).
52. Hagglund, P., Bunkenborg, J., Elortza, F., Jensen, O. N. & Roepstorff, P. A new strategy for identification of *N*-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J. Proteome Res.* **3**, 556–566 (2004).
53. Küster, B., Krogh, T. N., Mortz, E. & Harvey, D. J. Glycosylation analysis of gel-separated proteins. *Proteomics* **1**, 350–361 (2001).
54. Larsen, M. R., Hojrup, P. & Roepstorff, P. Characterization of gel-separated glycoproteins using two-step proteolytic digestion combined with sequential microcolumns and mass spectrometry. *Mol. Cell. Proteomics* **4**, 107–119 (2005).
55. McLachlin, D. T. & Chait, B. T. Improved β-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Anal. Chem.* **75**, 6826–6836 (2003).
56. Zhou, H., Watts, J. D. & Aebersold, R. A systematic approach to the analysis of protein phosphorylation. *Nature Biotechnol.* **19**, 375–378 (2001).
57. Tao, W. A. *et al.* Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. *Nature Methods* **2**, 591–598 (2005).
58. Brittain, S. M., Ficarro, S. B., Brock, A. & Peters, E. C. Enrichment and analysis of peptide subsets using fluorous affinity tags and mass spectrometry. *Nature Biotechnol.* **23**, 463–468 (2005).
59. Kho, Y. *et al.* A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *Proc. Natl Acad. Sci. USA* **101**, 12479–12484 (2004).
60. Sprung, R. *et al.* Tagging-via-substrate strategy for probing *O*-GlcNAc modified proteins. *J. Proteome Res.* **4**, 950–957 (2005).
61. Gevaert, K. *et al.* Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nature Biotechnol.* **21**, 566–569 (2003).
62. Van Damme, P. *et al.* Caspase-specific and nonspecific *in vivo* protein processing during Fas-induced apoptosis. *Nature Methods* **2**, 771–777 (2005).
63. Welchman, R. L., Gordon, C. & Mayer, R. J. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nature Rev. Mol. Cell Biol.* **6**, 599–609 (2005).
64. Giannakopoulos, N. V. *et al.* Proteomic identification of proteins conjugated to ISG15 in mouse and human cells. *Biochem. Biophys. Res. Commun.* **336**, 496–506 (2005).
65. Ghezzi, P. & Bonetto, V. Redox proteomics: identification of oxidatively modified proteins. *Proteomics* **3**, 1145–1153 (2003).

66. Aulak, K. S., Koeck, T., Crabb, J. W. & Stuehr, D. J. Proteomic method for identification of tyrosine-nitrated proteins. *Methods Mol. Biol.* **279**, 151–165 (2004).
67. Kanski, J. & Schoneich, C. Protein nitration in biological aging: proteomic and tandem mass spectrometric characterization of nitrated sites. *Methods Enzymol.* **396**, 160–171 (2005).
68. Miyagi, M. *et al.* Evidence that light modulates protein nitration in rat retina. *Mol. Cell. Proteomics* **1**, 293–303 (2002).
69. Kanski, J., Hong, S. J. & Schoneich, C. Proteomic analysis of protein nitration in aging skeletal muscle and identification of nitrotyrosine-containing sequences *in vivo* by nanoelectrospray ionization tandem mass spectrometry. *J. Biol. Chem.* **280**, 24261–24266 (2005).
70. MacCoss, M. J. *et al.* Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl Acad. Sci. USA* **99**, 7900–7905 (2002).
71. Wu, C. C., MacCoss, M. J., Howell, K. E. & Yates, J. R. 3rd. A method for the comprehensive proteomic analysis of membrane proteins. *Nature Biotechnol.* **21**, 532–538 (2003).
72. Elortza, F. *et al.* Proteomic analysis of glycosyl-phosphatidylinositol-anchored membrane proteins. *Mol. Cell. Proteomics* **2**, 1261–1270 (2003).
73. Steen, H., Jebanathirajah, J. A., Springer, M. & Kirschner, M. W. Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. *Proc. Natl Acad. Sci. USA* **102**, 3948–3953 (2005).
74. Oda, Y., Huang, K., Cross, F. R., Cowburn, D. & Chait, B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl Acad. Sci. USA* **96**, 6591–6596 (1999).
75. Heck, A. J. & Krijgsveld, J. Mass spectrometry-based quantitative proteomics. *Expert Rev. Proteomics* **1**, 317–326 (2004).
76. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl Acad. Sci. USA* **100**, 6940–6945 (2003).
77. Ross, P. L. *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–1169 (2004).
78. Krijgsveld, J. *et al.* Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nature Biotechnol.* **21**, 927–931 (2003).
79. Gruhler, A., Schulze, W. X., Matthiesen, R., Mann, M. & Jensen, O. N. Stable isotope labeling of *Arabidopsis thaliana* cells and quantitative proteomics by mass spectrometry. *Mol. Cell. Proteomics* **4**, 1697–1709 (2005).
80. Ibarrola, N., Molina, H., Iwahori, A. & Pandey, A. A novel proteomic approach for specific identification of tyrosine kinase substrates using [^{13}C] tyrosine. *J. Biol. Chem.* **279**, 15805–15813 (2004).
81. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
82. Zhu, H., Pan, S., Gu, S., Bradbury, E. M. & Chen, X. Amino acid residue specific stable isotope labeling for quantitative proteomics. *Rapid Commun. Mass Spectrom.* **16**, 2115–2123 (2002).
83. Kratchmarova, I., Blagoev, B., Haack-Sorensen, M., Kassem, M. & Mann, M. Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* **308**, 1472–1477 (2005).
84. Ballif, B. A. *et al.* Quantitative phosphorylation profiling of the ERK/p90 ribosomal S6 kinase-signaling cassette and its targets, the tuberous sclerosis tumor suppressors. *Proc. Natl Acad. Sci. USA* **102**, 667–672 (2005).
85. Guo, D. *et al.* A tethered catalysis, two-hybrid system to identify protein–protein interactions requiring post-translational modifications. *Nature Biotechnol.* **22**, 888–892 (2004).
86. Ptacek, J. *et al.* Global analysis of protein phosphorylation in yeast. *Nature* **438**, 679–684 (2005).
87. Gorg, A., Weiss, W. & Dunn, M. J. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**, 3665–3685 (2004). **An overview of 2D-gel electrophoresis, which includes its application to the study of protein diversity and PTMs.**
88. Ge, Y. *et al.* Multiplexed fluorescence detection of phosphorylation, glycosylation, and total protein in the proteomic analysis of breast cancer refractoriness. *Proteomics* **4**, 3464–3467 (2004).
89. Meng, F., Forbes, A. J., Miller, L. M. & Kelleher, N. L. Detection and localization of protein modifications by high resolution tandem mass spectrometry. *Mass Spectrom. Rev.* **24**, 126–134 (2005).
90. Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **40**, 430–443 (2005).
91. Uhlen, M. *et al.* A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* **4**, 1920–1932 (2005).
92. Kelleher, N. L. *et al.* Localization of labile posttranslational modifications by electron capture dissociation: the case of γ-carboxyglutamic acid. *Anal. Chem.* **71**, 4250–4253 (1999).
93. Zubarev, R. A. Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* **15**, 12–16 (2004).
94. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl Acad. Sci. USA* **101**, 9528–9533 (2004).
95. Forbes, A. J. *et al.* Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS. *Proc. Natl Acad. Sci. USA* **101**, 2678–2683 (2004).
96. Zabrouskov, V., Giacomelli, L., van Wijk, K. J. & McLafferty, F. W. A new approach for plant proteomics: characterization of chloroplast proteins of *Arabidopsis thaliana* by top-down mass spectrometry. *Mol. Cell. Proteomics* **2**, 1253–1260 (2003).
97. Jones, P. *et al.* PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res.* **34**, D659–D663 (2006).
98. Orchard, S., Hermjakob, H., Taylor, C., Aebersold, R. & Apweiler, R. Human proteome organisation proteomics standards initiative pre-congress initiative. *Proteomics* **5**, 4651–4652 (2005).
99. Peri, S. *et al.* Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* **13**, 2363–2371 (2003).
100. Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633–1649 (2004).
101. Chang, E. J., Archambault, V., McLachlin, D. T., Krutchinsky, A. N. & Chait, B. T. Analysis of protein phosphorylation by hypothesis-driven multiple-stage mass spectrometry. *Anal. Chem.* **76**, 4472–4483 (2004).
102. Hjerrild, M. *et al.* Identification of phosphorylation sites in protein kinase A substrates using artificial neural networks and mass spectrometry. *J. Proteome Res.* **3**, 426–433 (2004).
103. Nuhse, T. S., Stensballe, A., Jensen, O. N. & Peck, S. C. Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database. *Plant Cell* **16**, 2394–2405 (2004).
104. Schwartz, D. & Gygi, S. P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nature Biotechnol.* **23**, 1391–1398 (2005).
105. de Lichtenberg, U., Jensen, L. J., Brunak, S. & Bork, P. Dynamic complex formation during the yeast cell cycle. *Science* **307**, 724–727 (2005). **Shows the power of an integrative computational analysis of gene-expression data and protein–protein interaction data for the characterization of dynamic, functional protein modules.**
106. Rual, J. F. *et al.* Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173–1178 (2005).

Acknowledgements

I thank my colleagues in the Protein Research Group, as well as numerous colleagues in the world of proteomics, for many interesting and stimulating scientific discussions. I am grateful to M. B. Trelle for his assistance in preparing the figures. This manuscript is dedicated to the memory of my late father, Alex B. Jensen (1940–2005).

Competing interests statement

The author declares no competing financial interests.

DATABASES

The following terms in this article are linked online to:
UniProtKB: http://ca.expasy.org/sprot
ISG15 | NEDD8 | PNGase F | Ste20

FURTHER INFORMATION

Center for Biological Sequence Analysis Prediction Servers (Technical University of Denmark):
http://www.cbs.dtu.dk/services
Human Protein Reference Database (HPRD):
http://www.hprd.org
HUPO Proteomics Standards Initiative:
http://psidev.sourceforge.net
PhosphoELM (a database of S/T/Y phosphorylation sites):
http://phospho.elm.eu.org
Protein Research Group, University of Southern Denmark:
http://www.protein.sdu.dk
The Association of Biomolecular Resource Facilities:
http://www.abrf.org
UNIMOD (protein modifications for mass spectrometry):
http://www.unimod.org
Access to this links box is available online.