

Chapter 6

Post-translational Modifications in the Human Proteome

**Cheryl F. Lichti, Norelle C. Wildburger, Mark R. Emmett,
Ekaterina Mostovenko, Alexander S. Shavkunov,
Shinji K. Strain, and Carol L. Nilsson**

Abstract The Chromosome-Centric Human Proteome Project (C-HPP) is a global project aimed to identify at least one protein isoform encoded by the approximately 20, 300 human genes. In addition, protein post-translational modifications will be characterized, with the initial goal of detecting phosphorylation, acetylation, and glycosylation sites in each protein. In this chapter, we provide an overview of known post-translational modifications, their known biological functions, and present strategies to detect them on both a single protein and proteomic scales. In future proteomic studies, global characterization of post-translation modifications, splice variants, and variants caused by single nucleotide polymorphisms (SNPs) will be necessary to fully understand the role of proteins in human biology and disease.

Keywords Post-translational modification • Phosphorylation • Acetylation • Glycosylation • Ubiquitination • Mass spectrometry • C-HPP • Proteomics

C.F. Lichti • E. Mostovenko • A.S. Shavkunov • S.K. Strain • C.L. Nilsson (✉)
Department of Pharmacology, University of Texas Medical Branch,
301 University Blvd, Galveston, TX 77555, USA
e-mail: carol.nilsson@utmb.edu

N.C. Wildburger
Department of Pharmacology, University of Texas Medical Branch,
301 University Blvd, Galveston, TX 77555, USA

Department of Neuroscience and Cell Biology, University of Texas Medical Branch,
301 University Blvd, Galveston, TX 77555, USA

M.R. Emmett
Department of Biochemistry and Molecular Biology, University of Texas Medical Branch,
301 University Blvd, Galveston, TX 77555, USA

6.1 Introduction

The major advances in characterization of the human genome, including the first published sequence (Venter et al. 2001) and the first draft of the human genome “parts list” (Consortium 2012) promise to accelerate related projects that employ genome-wide analytical strategies. The Chromosome-centric Human Proteome Project (C-HPP) is a global research consortium that is charged with identification of all human proteins, defining their tissue and cellular expression, as well as mapping of the three major protein post-translational modifications (PTMs), acetylation, phosphorylation, and glycosylation. Because modifications change the structures and functions of proteins, systems-wide characterization of protein PTMs can significantly advance our knowledge of the role of PTMs in human development, health and disease.

The type and number of PTMs that have been characterized are highly diverse structurally and number over 100. The discoveries of new modifications are still being described in the scientific literature. While the goal of protein characterization within the C-HPP is limited to the description of protein acetylation, phosphorylation and glycosylation, there are other PTMs that can be captured on a proteomic scale, including modification by ubiquitin or small ubiquitin-like modifications (SUMO), lipidation, nitration, and even halogenation.

Protein PTMs are often dynamic, changing in response to intracellular or extracellular stimuli, to developmental signals or aging, in distinct tissue localizations or globally (Fig. 6.1). Furthermore, the interplay of systems of PTMs, including

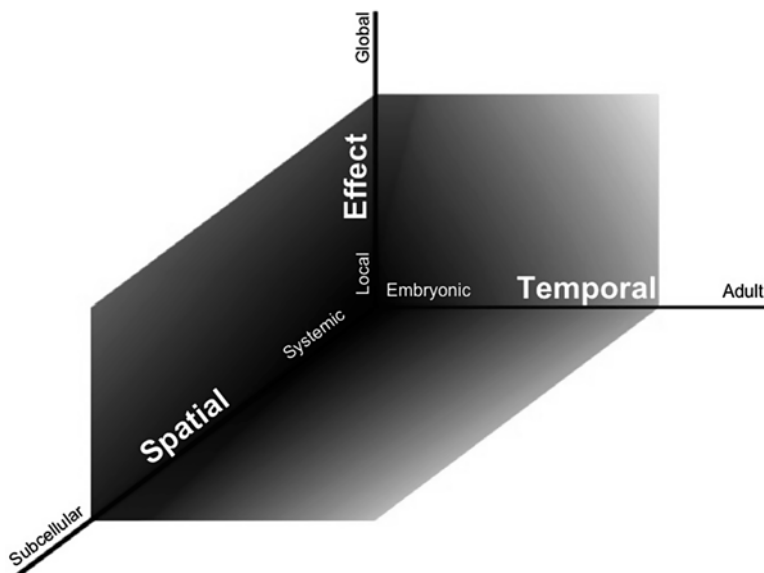


Fig. 6.1 Protein PTMs are dynamic, changing locally or globally in response to developmental signals, normal stimuli, or disease processes

Table 6.1 The protein families that control protein functions mediated by key PTMs. Examples of writers, readers and erasers as key control proteins in biological systems

Writers	Readers	Erasers
Histone acetylases	BRD proteins	Histone deacetylases
Kinases	SH ₂ -domain containing proteins	Phosphatases
Glycosyltransferases	Lectins	Glycosidases
Ubiquitin ligases	UB binding proteins	Deubiquitinases

The interplay of their activities contributes to the complexity of biochemical signaling pathways in health and disease states. Thus, integrated biological studies greatly benefit from defining phosphorylation, glycosylation and ubiquitin-mediated signaling

acetylation, phosphorylation, glycosylation and ubiquitination adds more layers to the complexity of signaling through those modifications. The fine regulation of protein PTMs is provided by key control proteins (Table 6.1), the so-called writers, readers and erasers of the chemical language embedded in PTMs.

Another characteristic of PTMs is that they are typically substoichiometric in proteolytic mixtures. Identification and site localization of PTMs often requires one or more enrichment steps. In this chapter, we describe functions associated with acetylation, phosphorylation, glycosylation and a few other selected PTMs. Further, analytical approaches to determine site localization and types of modifications on a proteomic scale are described.

6.2 Protein Acetylation

Proteins can be either stably N-terminally acetylated or reversibly acetylated on lysine (Lys) residues. Lys acetylation plays essential roles in cell homeostasis. A large number of studies indicate that reversible acetylation of Lys is widespread in the human proteome (Lin et al. 2014). The finding that nearly every enzyme in metabolic pathways, including glycolysis, has a plastic pattern of acetylation in response to factors such as nutritional status and disease, suggests that reversible acetylation is as important to the regulation of pathways as phosphorylation, glycosylation, and ubiquitination (Kouzarides 2000; Yuan and Marmorstein 2013).

The study of histone acetylation, induced by histone acetyltransferases (HATs), has been demonstrated to have a strong correlation to the regulation of gene activation (Eberharter and Becker 2002; Kimura et al. 2005; Yang and Seto 2007). Because the site localizations of histone PTMs are important to understand their function, mass spectrometry has played a prominent role in their analysis. Histones are smaller (~20 kDa) proteins, have heterogeneous modifications, and abundant basic amino acid residues. Because the proteolytic enzyme of choice in proteomics is trypsin, which cleaves at Lys and arginine (Arg) residues, histone digests yield an abundance of fragments with variable modifications. While site localization of PTMs in the tryptic peptide derived from a complex mixture of isoforms is feasible, it is not possible to assign PTMs reliably along the full length proteins.

Fortunately, sequencing by top-down mass spectrometry of smaller proteins is ideally suited for histone studies (Siuti and Kelleher 2007; Tipton et al. 2011). In top-down mass spectrometry, multiply charged gas phase protein ions are analyzed in a Fourier transform ion cyclotron resonance mass spectrometer (Marshall and Hendrickson 2008) and fragmented by electron capture dissociation (ECD) (Kelleher et al. 1999; Zubarev et al. 2000; Zubarev et al. 1998), infrared multiphoton dissociation (IRMPD) (Little et al. 1994), or a combination of both methods (Horn et al. 2000). The advantages of ECD sequencing of large polypeptides are that site localization of PTMs is easily achieved, even for labile modifications such as phosphorylation and O-glycosylation, and that those PTMs are assigned reliably for the entire protein isoform that is analyzed as an intact molecular ion. Furthermore, the high resolution and high mass accuracy that is characteristic of FT-ICR MS allows distinction between the nearly isobaric modifications acetylation and tri-methylation ($\Delta m = 0.0364$ Da); both of those PTMs may occur on histones, along with phosphorylation. The study of histone modifications, named epigenomics, is a very active area of scientific inquiry (Rivera and Ren 2013).

Some of the enzymes identified as HATs have also been shown to acetylate non-histone proteins, but most of the protein acetyltransferases acetylate non-histone proteins exclusively (Glozak et al. 2005; Spange et al. 2009). Even though acetylation is readily detected in mass spectrometry-based assays as a mass increase of 42.01 Da of a peptide or protein and a diagnostic immonium ion at m/z 126.1, the number of studies of the acetylome has lagged behind those aimed to map phosphorylation and ubiquitination sites. With the revelation that reversible Lys modification by acetylation plays a large role in cellular processes, the number of reports has increased greatly.

Reversible acetylation has been reported for transcription factors and may serve as an activating or inactivating modification (analogous to the role of phosphorylation). In the case of cellular tumor antigen p53 (TP53), polyacetylation leads to DNA binding, transcriptional activity and apoptotic functions (Luo et al. 2004; Sykes et al. 2006), whereas deacetylation represses the transcriptional activity by TP53 (Luo et al. 2000; Murphy et al. 1999). Reversible acetylation also plays an essential role in DNA replication, the physical separation of chromatids during mitosis (Heidinger-Pauli et al. 2009), and traffic of RNA through the nuclear pore complex (Bannister et al. 2000; Wang et al. 2004).

N-terminal acetylation of human proteins is a widespread (roughly 80 %) co-translational modification catalyzed by N-terminal acetyltransferases (NATs). Long viewed as a mere chemical block to protein degradation, it was recently discovered that N-acetylation of proteins can act as a signal for degradation by a ubiquitin ligase (Hwang et al. 2010). Furthermore, N-terminal acetylation can serve as a signal, an alternative to protein lipidation, for subcellular localization. Acetylation of ADP-ribosylation factor-like protein 8B (ARL8B) by NatC targets the protein to the lysosomal membrane (Starheim et al. 2009). Aberrant N-terminal acetylation of hemoglobin can cause clinical diseases related to resulting abnormal oxygen-binding capacity of the holoprotein (Starheim et al. 2009; Manning et al. 2012).

Like other modified peptides, acetylated peptides may be difficult to isolate and sequence by tandem mass spectrometry (MS/MS) when they are analyzed in the background of a complex mixture. Global studies of N-terminal acetylation have been enabled by special enrichment techniques prior to analysis by liquid chromatography (LC)-MS/MS. Gevaert et al. devised a negative enrichment scheme in tryptic digests of a proteome are reacted with 2,4,6-trinitrobenzenesulfonic acid. The reactive agent couples to the N-terminus of internal tryptic peptides. Upon separation of the complex mixture by C₁₈-based LC, the trinitrophenyl-bearing internal peptides are easily separated from the N-terminally acetylated peptides (Van Damme et al. 2011). Another approach that has met with success is to perform pre-separation of complex proteolytic digests by strong cation exchange chromatography (SCX) at low pH (Van Damme et al. 2011; Dormeyer et al. 2007; Crimmins et al. 1988).

6.3 Protein Phosphorylation

O-phosphorylation of protein serine (Ser), threonine (Thr), and tyrosine (Tyr) residues, a reversible process, is a well-studied mechanism of cell signaling and its regulation. Phosphorylation mainly occurs on Ser and Thr, just a few percent of phosphorylation occurs on Tyr. Protein phosphorylation is mediated by protein kinases, a superfamily of roughly 500 proteins which occupy approximately 3 % of the human genome (Manning et al. 2002). Protein phosphorylation may either increase or decrease protein activity (acting as an on-off switch), trigger a change in protein-protein binding characteristics or subcellular localization. For instance, phosphorylation of transcription factors often causes dimerization and translocation into the cell nucleus. Signal transducer and activator of transcription 3 (STAT3), a multifunctional transcription factor is archetypical in this respect (Reich 2009). The regulation of signaling pathways by protein phosphorylation has been intensively studied, especially for phospho-Tyr-mediated pathways. Historically, the bias may largely be due to the availability of reliable phospho-Tyr antibodies; antibodies for phospho-Ser and phospho-Thr antibodies typically have poorer specificity for the intended antigen. With newer, MS based sequencing strategies, thousands of phosphorylation sites can be assigned in a single experiment (Oppermann et al. 2009). Derivation of the biological meaning of changes in Ser and Thr phosphorylation can be quite challenging because relatively little has been reported in the literature on this topic.

Protein phosphorylation is typically substoichiometric, heterogeneous with respect site localization, and transitory in nature. These characteristics make MS an important tool for characterization of both single protein phosphorylation and global phosphoproteomics, because in contrast to antibody-based methods, data is acquired at both high sensitivity and specificity (Nilsson 2011a). However, there are significant challenges even in MS-based approaches. Phosphopeptides are substoichiometric and usually require an enrichment step. They are acidic in nature and

thus ionize poorly in positive ion mode, in the presence of unmodified peptides (ion suppression). Furthermore, Ser- and Thr-phosphorylation are thermolabile and may dissociate with prompt loss of the phosphate group, making peptide identification and site localization of the phosphorylation site more difficult.

Sample preparation alternatives for phosphoproteomics are important to consider at the experimental planning stage. While mainly chromatographic methods are employed in LC-tandem mass spectrometry (MS/MS) workflows, gel-based methods may also be applied (Eyrich et al. 2011; Dephoure et al. 2013; Černý et al. 2013). Anion exchange methods for peptides with acidic modifications (phosphate, sialic acid, and acetylation) are particularly widespread in the literature. In SCX, a column is packed with anionic resin. A salt gradient in the mobile phases induces elution of analytes based on increasing isoelectric point. When protein or peptide mixtures pass over the column, more basic polymers are retained longer, while more acidic compounds elute off the column in the earlier fractions (Gilar et al. 2005). Thus SCX, despite its inherent low peak capacity, may benefit experiments in which enrichment of phosphorylated, acetylated or sialylated peptides is desirable (Gilar et al. 2008).

Hydrophilic interaction chromatography (HILIC) (Alpert 1990) is a separation mode that fractionates peptides or proteins based on their polarity. The approach employs a polar resin and a partly hydrophilic mobile phase. Unlike reversed phase (RP) chromatography which retains analytes based on hydrophobicity, HILIC retention is higher for hydrophilic compounds. This is important because highly hydrophilic peptides (phosphorylated and glycosylated) may not be retained at all on RP (C_{18} -based) resins, but wash off in the flow through part of the gradient. Electrostatic repulsion hydrophilic interaction chromatography or ERLIC was introduced by Alpert (2008). ERLIC is a derivation of HILIC and allows the isocratic separation of phosphopeptides and glycopeptides from a proteome digest. In ERLIC, the column matrix and analytes share the same charge resulting in electrostatic repulsion, yet the mobile phase contains enough organic solvent to force the analytes to remain on the column through hydrophilic interactions (Alpert 2008). The carboxylic acids of the C-termini, and acidic amino acid (aspartate and glutamate) side chains become protonated ($-\text{COOH}$) in low pH conditions (Mysling et al. 2010). These neutral ion-pairs have much higher hydrophobicity than their charged state (Ding et al. 2007; Wimley et al. 1996).

Metal-based enrichment depends on immobilized metal cations which can bind Lewis base (electron pair donating) groups on phosphates and sialic acids (Larsen et al. 2005). Immobilized metal affinity chromatography (IMAC) is widely used and requires binding of metal ions (such as Fe^{3+} , Ti^{4+} , or Zr^{4+}) to a solid surface or particle. The metal oxide approach (MOAC) is similar in nature as it requires binding of a metal oxide (TiO_2 , ZrO_2) to a matrix material. The results obtained from commercially obtained MOAC media are quite varied and testing is recommended prior to large-scale studies using this enrichment approach. For reviews on this subject, two recent papers are recommended for further reading (Gates et al. 2010; Ficarro et al. 2009). IMAC is the most widely applied technique for phosphopeptide analysis, based on the number of literature references.

6.4 Protein Glycosylation

Glycosylation is necessary for protein folding, solubility, stability, trafficking, cell-cell communication, and adhesion (Varki 1993; Imperiali and Rickert 1995). It is estimated that approximately 50 % of all proteins are glycosylated, though only approximately 10 % of known proteins have been annotated as such (Apweiler et al. 1999). Unlike the proteins they modify, glycosylation synthesis is non-template driven, which produces highly complex glycan structures with large variations in branching points, linkage, monosaccharide composition, and configuration (Dell and Morris 2001). As such, it is the only PTM that requires detailed structural characterization.

There are two main types of protein glycosylation, N-linked glycosylation where the oligosaccharide is covalently attached to an asparagine (Asn) and O-linked glycosylation in which the attachment occurs on the hydroxyl group of either serine (Ser) or threonine (Thr). Protein glycosylation is tightly regulated in a series of enzymatic steps and its contribution to protein function is variable. In some instances, the lack of N-linked glycosylation targets the nascent polypeptide for degradation, while in other cases protein folding and secretion remains mildly affected if at all. N-glycans are synthesized in the endoplasmic reticulum (ER) on a dolichol donor and transferred en bloc onto nascent proteins co-translationally (Nilsson and von Heijne 1993). The first monosaccharide, N-acetylglucosamine (GlcNAc), of the trimannosyl-chitobiose core common to all N-glycans is linked via an amide bond to the asparagine (Asn) (Fig. 6.2) within the consensus sequon Asn-X-Ser/Thr (where X is any amino acid except proline) and occasionally Asn-X-Cys (Satomi et al. 2004). After attachment of the N-glycan complex, the nascent polypeptide chain enters the calnexin/calreticulin pathway, whereby the glycan complex acts as a ligand to calnexin and calreticulin, which sequesters the nascent glycopolypeptide chain for proper protein folding. Properly folded proteins leave the ER for the Golgi where the carbohydrate moiety is modified by glycosyltransferases and glycosidases (Dell and Morris 2001). O-linked glycosylation, unlike N-linked, occurs in the Golgi rather than the ER with the exception of O-mannosylation. In contrast to N-linked glycosylation, there is no consensus sequon for O-linked glycosylation, thus any Ser or Thr residue is a potential O-glycosylation site.

The close connection between changes in glycosylation and the development of cancer is well documented (Dube and Bertozzi 2005). Glycans may be over- or underexpressed compared to normal tissues and reappearance of embryonic types of glycans may re-emerge. Further, many human disorders are related to congenital disorders of glycan synthesis or degradation (CDG). While relatively rare, they serve as a reminder of the importance of glycosylation to normal function. In alpha-mannosidosis, the molecular deficit is inactivity of the enzyme responsible for hydrolyzing mannose that is alpha-linked in N-linked glycans. The clinical symptoms are defects in the immune system, abnormalities in skeletal development, hearing impairment, and mental retardation (Malm and Nilssen 2008). CDGs can also be related to deficiencies in protein sialylation, leading to severe morphogenic

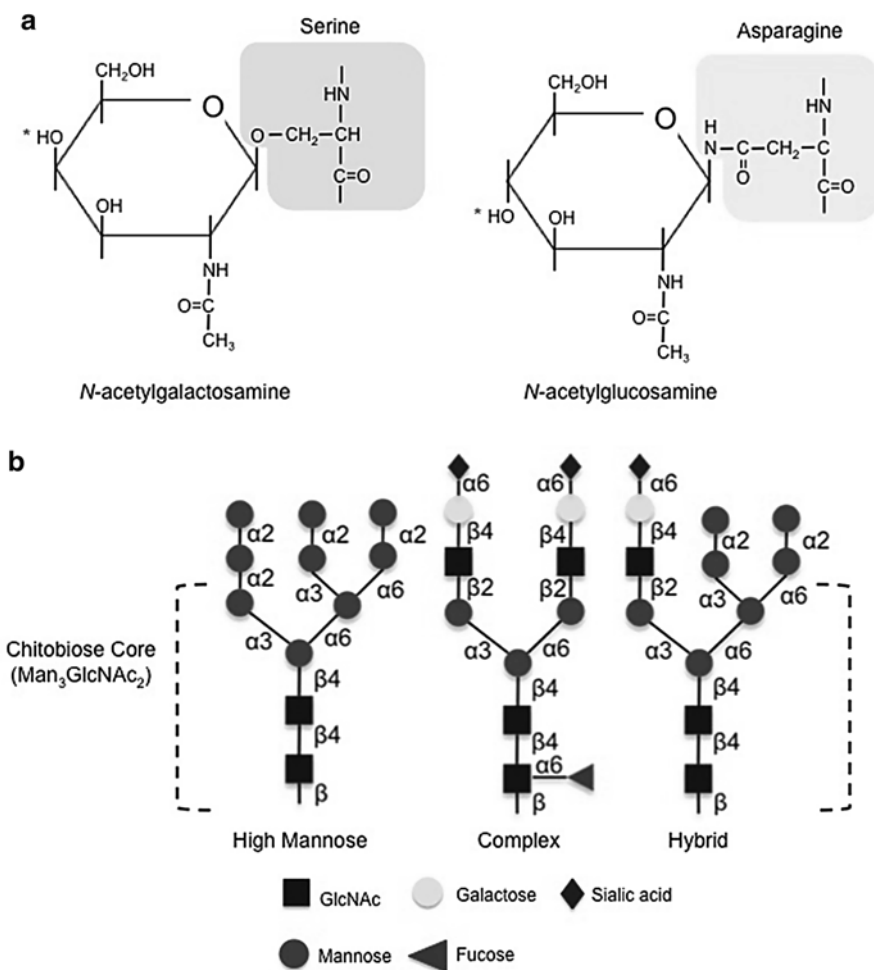


Fig. 6.2 (a) Structure of O- and N-linked glycan attachment to peptide (*left* and *right*, respectively). (b) Types of N-linked glycans

and metabolic abnormalities as well as shortened lifespan. The serum protein transferrin is normally a highly sialylated protein, lower than normal sialic acid content of this abundant protein is considered to be a biomarker for CDG (Schachter and Freeze 2009). Other defects have been linked to congenital muscular dystrophy syndromes (Martin-Rendon and Blake 2003).

Protein glycosylation is not static; its context is both spatially and temporally dependent. Though less well studied, evidence exists for the differential spatial distribution of glycoproteins. Tenascin-R, a neural extracellular matrix protein predominantly expressed in the cerebellum carries a terminal GalNAc-4-SO₄ on its

N-linked oligosaccharides (GalNAc; N-acetylgalactosamine). However within the cerebellum, only Purkinje cell bodies and their dendrites express GalNAc-4-SO₄ modified tenascin-R (Woodworth et al. 2002).

Temporally, poly(α 2,8)sialic acid (PSA) is highly abundant in the brain during the early stages of development, but gradually decreases over time (Varki et al. 2009). PSA is a highly anionic homopolymer of up to 300 α 2,8-linked sialic acids (Varki et al. 2009). PSA-modified neural cell-adhesion molecule at synapses reduces cell-cell and cell-matrix/extracellular adhesion likely through electrostatic repulsion by the multitude of anionic charges and large space created by the hydration volume (Varki et al. 2009), the result is a global reduction in membrane-membrane contact, effecting cell-to-cell contact with ligands, receptors, and adhesion molecules, in early development (Rutishauser 2008).

Local effects of protein glycosylation (where the precise location of the modification matters) are equally important. The receptor tyrosine kinase EPHA2 ligand Ephrin-A1 (EFNA1) contains one consensus sequon within its amino acid sequence at N26. Crystallography of the EPHA2/EFNA1 receptor-ligand complex confirmed glycosylation at N26 (Himanen et al. 2009; Himanen et al. 2010), and removal of glycosylation at this site results in the loss of EFNA1 biological function (Ferluga et al. 2013). The carbohydrate moiety of EFNA1 is essential to interact with the binding domain of EPHA2 and stabilizes the ligand-receptor interaction and subsequent tetramerization (Ferluga et al. 2013).

O-GlcNAc and phosphorylation are two different PTMs that may interact competitively, reciprocally, or simultaneously (Zeidan and Hart 2010; Hart et al. 2011). Most tumor suppressor proteins are modified by O-GlcNAc. Myc proto-oncogene protein Myc (MYC) is phosphorylated at Thr58 within the N-terminal transcriptional activation domain, which is required for MYC-dependent gene activity (Gupta et al. 1993). However, Thr58 is also modified by O-GlcNAc. Whereas phosphorylation of Thr58 promotes c-Myc transcriptional activity, the addition of O-GlcNAc likely attenuates its activity as growth inhibited cells contain predominantly O-GlcNAc modified MYC (Zeidan and Hart 2010). In another example, the addition of O-GlcNAc at Ser149 on TP53 inhibits phosphorylation at Thr155 thereby indirectly preventing ubiquitination (Yang et al. 2006). TP53 is just one example of proteins whose functions are regulated by the combination of acetylation, phosphorylation, glycosylation, and ubiquitination, underscoring the importance of mapping PTMs in the C-HPP experiments.

Glycoproteins and peptides present an analytical challenge as they are substoichiometric relative to non-glycosylated peptides, may have only partial sites of occupancy, and ionize poorly compared to their non-modified peptides (Dell and Morris 2001). These factors make it necessary to separate glycopeptides from peptides through enrichment strategies. A number of methods exist such as: lectins (Bunkenborg et al. 2004; Nilsson 2011b), hydrazide chemistry (Zhang and Aebersold 2006), graphitized carbon (Davies et al. 1992; Larsen et al. 2005), titanium dioxide (Larsen et al. 2007), strong cation exchange (Lewandrowski et al. 2007), HILIC (Hägglund et al. 2004; Wührer et al. 2004; Boersema et al. 2008; Di Palma et al. 2012), and ERLIC (Alpert 2007).

Lectin affinity chromatography can be used to enrich either glycoproteins or glycopeptides, but is more robust at the glycoprotein level (Atwood et al. 2006). On the other hand, highly abundant non-glycosylated proteins tend to cause 'carry over' in the glycoprotein fraction during lectin chromatography (Atwood et al. 2006; Bunkenborg et al. 2004). Lectin affinity purification at the peptide level minimizes this effect. Additional experimental considerations include the broad specificities of different lectins and the nature of the sample itself. For instance, membrane proteins require detergents or chaotropes to remain soluble, yet these reagents interfere with lectin binding; therefore, a digestion step prior to lectin enrichment is recommended.

Hydrazide chemistry captures glycoproteins on hydrazide resin after oxidation of the carbohydrate moieties (Zhang et al. 2003). Protein digestion with trypsin followed by extensive washing removes unmodified peptides. At this stage, isotopic labeling of glycopeptides for quantification occurs before removal from the hydrazide resin by PNGase F cleavage (see below). Hydrazide enrichment has limitations, including not being amenable to automation, requiring extensive reactions with toxic chemicals, prolonged enrichment, and extensive sample cleanup.

Graphitized carbon cartridges (GCC), while predominately used for the isomeric separation (Koizumi et al. 1991) of free glycans, is also applicable to the enrichment and separation of glycopeptides (Davies et al. 1992; Fan et al. 1994). GCC itself is more chemically and physically stable than silica and can be used in a broad pH range from highly acidic to highly alkaline. The effect of temperature is not nearly as drastic as it is on silica-based resins and GCC can remove salts and detergents prior to mass spectrometry analysis (Packer et al. 1998).

Tryptic peptides are usually too large for either enrichment (offline SPE) or separation (online nLC-MS/MS) of glycopeptides. However, GCC is well suited for smaller peptides resulting from a multienzyme or pronase digest. Pronase leads to nearly complete digestion into individual amino acids. In the case of glycopeptides, the glycan structure prevents complete enzymatic cleavage due to steric hindrance (in the case of N-linked glycopeptides) (An et al. 2003; Hua et al. 2013), resulting in a carbohydrate moiety attached to a peptide backbone up to eight amino acids long. One shortcoming of this strategy is that the peptide length may not always be optimal for database searching and protein identification.

Titanium dioxide (TiO_2) and SCX were originally applied to improve the enrichment of phosphopeptides. In recent years, Larsen and coinvestigators as well as Sickmann and colleagues adapted TiO_2 and SCX, respectively, for enrichment of terminal sialic acid containing N-linked glycopeptides (Larsen et al. 2007; Palmisano et al. 2011; Lewandrowski et al. 2007). TiO_2 (and ZrO_2)-based enrichment depends on immobilized metal cations. TiO_2 retains negatively charged sialoglycopeptides through multiple interactions with the hydroxyl and carboxyl groups (Larsen et al. 2005). However, TiO_2 also enriches phosphopeptides and peptides rich in acidic residues due to their negative charge, making it necessary to treat samples with phosphatase to prevent oversaturation of the column and improve separations. In contrast to TiO_2 , sialoglycopeptides elute early in SCX fractionation. The negative charge of sialic acid counteracts the overall positive charge of the peptide (in low pH

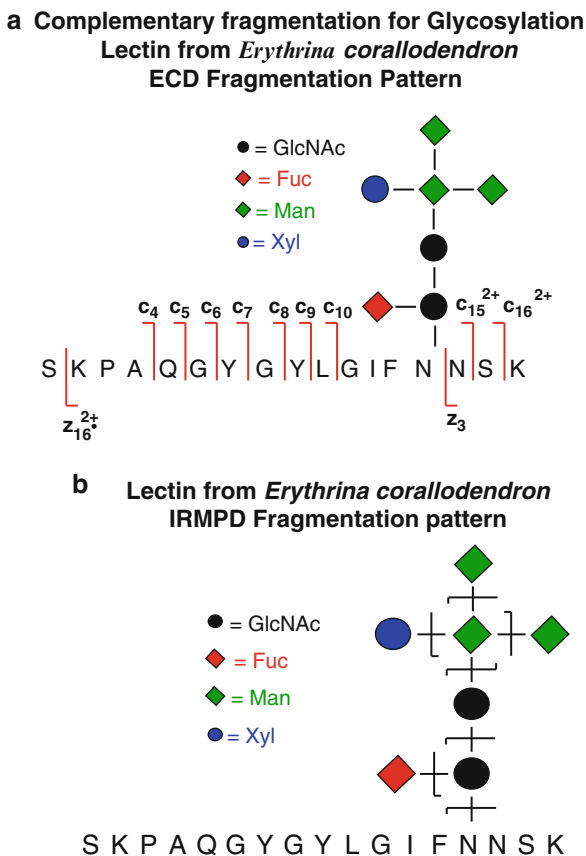
solutions) resulting in little net charge. The sialoglycopeptides carrying relatively little net charge elute in earlier fractions compared to unmodified peptides (Lewandrowski et al. 2007). SCX also enriches peptides with other acidic modifications such as phosphorylation and acetylation. Similarly to TiO_2 , phosphatase treatment partially remedies this issue.

Hydrophilic interaction liquid chromatography (HILIC) is another enrichment technique to separate the glycopeptides from peptides. Glycopeptides are generally more hydrophilic than non-glycosylated peptides, which allows retention on the stationary phase through hydrophilic partitioning, based on hydrogen bonding and to some extent electrostatic interactions depending on the type of stationary phase used (Alpert 1990). However, there exists a hydrophilic overlap between non-glycopeptides and glycopeptides (Mysling et al. 2010). In complex samples, this becomes readily apparent when these peptides co-elute with glycopeptides and result in ion suppression of the glycopeptides of interest during MS analysis. The use of an ion pairing agent in the solvent system can improve the separation.

Electrostatic repulsion hydrophilic interaction chromatography or ERLIC was introduced by Alpert (Alpert 2008). The carboxylic acids of the C-termini, and acidic amino acid (aspartate and glutamate) side chains become protonated ($-\text{COOH}$) in low pH conditions (Mysling et al. 2010). These neutral ion-pairs have much higher hydrophobicity than their charged state (Ding et al. 2007; Wimley et al. 1996). The hydrophilic glycans ($-\text{OH}$ groups) of the glycopeptides remain unaffected by the ion-pairing agent because of the non-ionic electrostatic interactions between the carbohydrates and stationary phase (Ding et al. 2007; Wimley et al. 1996). ERLIC, a derivation of HILIC (i.e., anion-exchange HILIC) allows the isocratic separation of phosphopeptides and glycopeptides from a tryptic proteome digest. In ERLIC, the column matrix and analytes share the same charge resulting in electrostatic repulsion, yet the mobile phase contains enough organic solvent to force the analytes to remain on the column through hydrophilic interactions (Alpert 2008). Recently several groups demonstrated the application of ERLIC for the simultaneous enrichment of glyco- and phosphopeptides in the same sample in a single chromatographic run (Zhang et al. 2010; Hao et al. 2011).

Peptide N-glycosidase F (PNGase F) hydrolyzes the β -aspartylglycosylamine bond linking the first GlcNAc of the core to the nitrogen of asparagine (N) for all N-linked glycans (except in cases of core α 1,3-fucosylation). This converts the asparagine through a deamidation reaction to aspartic acid (D). Deamidation resulting from PNGase F deglycosylation yields a mass increase of +0.98 Da. This facilitates identification of N-linked glycosylation sites (when analyzed with high-resolution MS), but cannot completely guard against false positives due to spontaneous deamidation within a consensus sequence, even with isotopic (^{18}O) labeling (Robinson and Robinson 2001; Palmisano et al. 2012). This makes it absolutely necessary to have identical control samples or aliquots of the same sample without PNGase F treatment. Since glycosylation suppresses peptide ionization, it would be unlikely to see the same peptide by MS/MS if it were glycosylated. Conversely, if identified by MS/MS in the control sample, then the deamidation within the consensus sequon is most likely an artifact and a 'false positive'. 'True

Fig. 6.3 The high degree of complementarity of ECD (a) and IRMPD (b) to derive structural data from N-linked glycopeptides



positive’ sites may only be annotated as “putative’ as they are observed only after enzymatic deglycosylation. An alternative strategy would be to keep the glycan moiety intact on its peptide backbone and analyze by multistage MS in an ion trap instrument (Reinhold et al. 2013), or by a combination of fragmentation techniques that fragment the glycan and the peptide backbone separately (e.g., IRMPD/ECD, CID/ETD, CID/HCD, or HCD/ETD) (Hakansson et al. 2001; Wu et al. 2007; Alley et al. 2009; Scott et al. 2011; Singh et al. 2012). Such strategies can provide complementary structural datasets (Fig. 6.3).

Alternatively, other available enzymes such as Endo H and the Endo F family (I-III) cleave within the chitobiose core between the two GlcNAc residues, leaving a single GlcNAc or fucosylated GlcNAc on the peptide. The mono- or disaccharide containing peptide can then be fragmented with MS/MS. If CID is employed, GlcNAc may be lost from the peptide backbone as an oxonium ion of 203.08 Da (or 349.14 Da in the case of fucosylated GlcNAc). Provided there is only one N-linked sequon in the peptide, site occupancy can be determined with the presence of the diagnostic oxonium ion (Hägglund et al. 2004, 2007).

In general, the same strategies described above for the analysis of N-linked glycoproteins can be applied to O-linked glycopeptides, with the exception that there is no universal enzyme, like PNGase F, for cleaving O-linked glycans. O-linked glycosylation (i.e., GalNAc) is typically shorter in length than N-linked glycosylation. Yet in many cases just as complex due to extended cores. For O-GlcNAc, due to its extreme thermolability, ETD or ECD fragmentation is recommended to determine site occupancy.

The remarkable diversity and molecular choreography of glycosylation within a spatiotemporal biological context allows cells to fine-tune the biological and biophysical properties of proteins, vastly expanding molecular communication and functional outcomes (Edwards et al. 2014). With the ever-increasing knowledge of protein glycosylation and other PTMs, the next major and necessary task will be the large-scale elucidation of PTM structural-functional relationships.

6.5 Protein Ubiquitination

Ubiquitin (Ub) is a small protein, consisting of 76 amino acids, whose reversible covalent attachment to proteins governs such diverse biological processes as proteasomal degradation, DNA repair, activation of transcription factors, intracellular trafficking, and regulation of histones. Four different human genes are known to code for ubiquitin: UBB and UBC, which both code for polyubiquitin chains, and UBA52 and RPS27C, which code for single copies of ubiquitin fused to L40 and S27A, respectively.

In addition to Ub, several other small, ubiquitin-like proteins (Ubls) can be similarly attached to the lysine residue of a target protein. These Ubls include three isoforms of small ubiquitin-related modifier (SUMO1, SUMO2, SUMO3), neural precursor cell expressed, developmentally down-regulated 8 (NEDD8), ubiquitin-like protein ISG15 (ISG15), ubiquitin D (UBD), ubiquitin-like 5 (UBL5), ubiquitin-related modifier 1 homolog (URM1), autophagy associated protein 8 (ATG8), and ubiquitin-like protein ATG12 (ATG12). The focus of this discussion will be Ub, but occasional comparisons will be made to other Ubls.

Ub is attached to proteins through its C-terminal glycine residue via the epsilon amino group of a Lys residue in the target protein, forming an amide bond, through a series of enzymatic reactions (Fig. 6.4). In the first of these reactions, Ub is bound by an Ub-activating enzyme (E1), along with Mg^{2+} and ATP. E1 then catalyzes C-terminal acyl adenylation of the Ub chain, activating it for further reaction with a cysteine sulfhydryl group of E1. The activated Ub is then transferred to an Ub-conjugating enzyme, E2, through a transthioesterification reaction. Attachment to the target protein occurs through ubiquitin ligases, E3. The attachment of Ubls follows the same pathway, with characteristic E1/E2/E3 enzymes to catalyze the process. For both Ub and Ubls, the species attached can either be monomeric or polymeric, and the nature of the attachment helps signal the functional role of the modification.

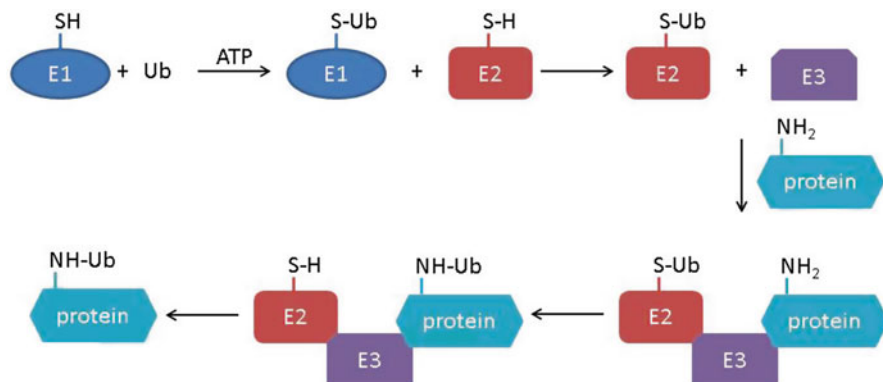


Fig. 6.4 The synthetic pathway involved in protein ubiquitination. Ub is attached to a Cys sulfhydryl residue of a Ub-activating enzyme (E1) in a process that is catalyzed by Mg^{2+} and ATP. The activated Ub is then transferred to the Cys sulfhydryl group of an Ub-conjugating enzyme (E2). Attachment to the target protein is mediated by a ubiquitin ligase (E3), which recruits both the target protein and an Ub-charged E2

Ubiquitin itself has seven Lys residues (at positions 6, 11, 27, 29, 33, 48 and 63) in addition to the N-terminus, and all provide potential sites for the Ub chain to be expanded. The signal that is sent by polyubiquitination is determined by which of these residues is modified. PolyUb chains can be linear, formed through end-to-end linkages via the N-terminus, or they can be branched through attachment at any of the other Lys residues. Attachment of at least four Ubs at Lys 48 is a known signal for proteasomal degradation (Voutsadakis 2007), and Ub chains at Lys 6 and Lys 11 have also been seen as proteasomal degradation signals (Voutsadakis 2012). Lys 63 ubiquitination can be a signal for autophagy, DNA repair or receptor kinase endocytosis (Jadhav and Wooten 2009).

To date, there are eight known E1 enzymes for activation of Ub and UbIs. (Schulman and Harper 2009). Interestingly, there are two known E1s for ubiquitin (UBA1 and UBA6), while a single E1 (SAE1-UBA2 heterodimer) is responsible for the activation of all SUMOs. Although each E1 recognizes Ub or a particular class of Ubl, all share a common mechanism of action. Key to the action is the adenylation domain, which can be homo- or heterodimeric in various members of the E1 family. In addition to catalyzing the acyl adenylation reaction, this domain's key function is to recognize the Ub or Ubl whose reaction it is catalyzing. Once the Ub, ATP and Mg^{2+} are positioned in the active site, the C-terminus is acyl adenylated, liberating inorganic phosphate. The Ub acyl adenylate is then attacked by a cysteine sulfhydryl group from the catalytic cysteine domain, resulting in formation of a new thioester linkage between Ub and the cysteine sulfhydryl group. Once this happens, a second Ub is then acyl adenylated and bound in the adenylation domain. The role of this second Ub is not known but is postulated to play a role in stabilizing the active conformation of the enzyme for transferring Ub between E1 and E2. Formation of the E1-Ub thioester triggers association of the appropriate E2 enzyme

and subsequent Ub transfer. The ubiquitin fold domain (UFD) plays a key role in recognizing and binding to the correct E2, and a conformational change in the UFD is necessary to bring the E1-thioubiquitin close to the reactive E2 cysteine residue so that transthioesterification can occur.

Approximately forty ubiquitin E2s are encoded in the human genome, and all possess a highly conserved region of 150–200 amino acids that is known as the ubiquitin-conjugating catalytic (UBC) fold (van Wijk and Timmers 2010). The UBC is of central importance in the function of E2s as it is involved in binding E1s, E3s, and the activated Ub. Classification of E2s is determined by the presence or absence of additional sequence modification to the UBC. Class I E2s contain only the UBC fold, while Class II and III E2s contain additional sequence on the N- and C-termini, respectively. Class IV E2s are modified at both the N- and C-termini. The differences in sequence between the four classes of E2s help to define their differences in subcellular localization, interaction with E1s and E3s, and modulation of the activity of an interacting E3 (van Wijk and Timmers 2010). However, the main determinants for E2 properties are found within key regions of the UBC. An anti-parallel beta sheet and one alpha helix (H2) form a central region, bounded by alpha helix 1 (H1) on one side and alpha helices 3 and 4 (H3 and H4) on the other.

Substrate specificity and recruitment are mediated by E3 ubiquitin ligases, and the fact that there are ~600 putative E3 ligases encoded into the human genome provides clues regarding the diversity of substrates for ubiquitination. As with E2s, E3s can be either monomeric, homodimeric or heterodimeric. E3s fall into two broad classes, based upon the domain responsible for binding to E2. Most human E3s contain a RING (Really Interesting New Gene) domain (Metzger et al. 2014). Two Zn^{2+} ions are bound in the RING domain, providing a scaffold for binding E2. A small subset of this first class of E3s, the RING-like U-box domain family, binds E2 in a similar manner but without the requirement for Zn^{2+} ions. The RING and RING-like E3s mediate the transfer of ubiquitin from E2 to the substrate of interest without themselves being modified by ubiquitin, instead facilitating the process by bringing the reactive species in close proximity to one another. In contrast, E3s in the second category are modified by ubiquitin in an intermediate catalytic step. This family of E3s contains a HECT (Homologous to E6-AP Carboxy Terminus) domain, and a conserved cysteine residue serves to transfer Ub to the substrate via formation of an intermediate E3-Ub thioester (Scheffner and Kumar 2014). In the case of the HECT family of EC ligases, Ub or Ubl specificity is found within the catalytic HECT domain. For RING family E3s, the Ub or Ubl whose attachment is being catalyzed depends upon the specific combination of E2 and E3. In other words, RING family E3s can generate different Ub linkages depending upon the E2 to which it is interacting. For HECT family E3s, protein substrate specificity is governed by a region that is located on the N-terminal side of the HECT domain.

Factors governing substrate recognition are not completely understood, but some trends are beginning to emerge (Jadhav and Wooten 2009). For instance, it is thought that cytosolic E3s recognize misfolded proteins through long stretches of hydrophobic residues. Other signals may include post-translational modifications or primary sequence. In terms of primary sequence, proteins with N-terminal Phe, Leu, Asp,

Lys, or Arg residues were found to have a very short half-life (2–3 min.) when compared to amino acids with “stabilizing” the N-terminal residues Met, Ser, Ala, Thr, Val and Gly (Bachmair et al. 1986). This is known as the N-end rule. The destabilizing amino acids plus the Lys residue to be modified with Ub are referred to as an N-degron; they are recognized by E3s called N-recognins and targeted for destruction by the 26S proteasome (Jadhav and Wooten 2009). Other important primary sequence signals for ubiquitination include PEST sequences, so called because the sequences are rich in the amino acids Pro, Glu, Ser and Thr (Rogers et al. 1986); D-box, containing the consensus sequence R-A/T-A-L-G-X-I/V-G/T-N (Glutzer et al. 1991); and the KEN (Lys-Glu-Asn) box domain, with the consensus sequence K-E-N-X-X-X-N (Pfleger and Kirschner 2000). Post-translational modifications, including phosphorylation and glycosylation, have also been shown to activate a substrate toward ubiquitination.

As is the case with many post-translational modifications, ubiquitination is a reversible reaction. The proteolytic cleavage of ubiquitin side chains is catalyzed by a class of enzymes known as deubiquitinases (DUBs). At present, there are approximately 79 known DUBs (Komander et al. 2009) which follow several mechanistic pathways depending upon the polyubiquitin bond being proteolytically cleaved. As previously mentioned, ubiquitin can be transcribed as a polyUb chain or as fusion protein, so a DUB is crucial in producing monoUb. Also, once proteins have been targeted for degradation, the Ub and polyUb chains can be removed by DUBs for the purpose of recycling Ub and maintaining a steady state intracellular Ub concentration. DUBs can also remove Ub and polyUb in order to reverse Ub signaling and rescue proteins from degradation. There are five families of DUBs, and all possess a Ub-binding domain (UBD). Most DUBs catalyze cleavage of the bond between the epsilon-amino group of Lys and the C-terminus of Ub.

As is the case with many PTMs, the substoichiometric nature of ubiquitination can lead to challenges in identification of modified proteins. An additional challenge lies in the fact that ubiquitinated proteins are typically targeted for degradation. Therefore, enrichment strategies are a key component in the global mapping of ubiquitination. One of the earliest reported strategies for global study of ubiquitination involved a cell culture study of a cell line expressing a His-tagged form of Ub. A Ni-NTA column was used to isolate proteins modified by Ub. Proteolytic digestion with trypsin and LC-MS/MS analysis yielded 4210 peptides corresponding to 1,075 candidate ubiquitin-conjugating proteins (Peng et al. 2003). Ub-binding domains, covalently attached to beads, have also been shown to be effective in the enrichment of Ub proteins (Nakayasu et al. 2013), as have linkage-specific polyubiquitin antibodies (Matsumoto et al. 2010; Newton et al. 2008).

When ubiquitinated peptides are proteolytically digested with trypsin, a Gly-Gly motif remains on ubiquitinated lysine residues. Also, due to the modification of the Lys, trypsin will not cleave at this residue. The mass of the Gly-Gly remnant is 114.04 Da, and database search engines can use this value to identify modified peptides. Some caution must be used, though, because this particular change in mass is also characteristic for the addition of two carbamidomethyl groups, formed from over-alkylation with iodoacetamide. The Gly-Gly tag is also formed upon cleavage

of two Ub-like proteins (NEDD8 and ISG15), so care must be taken when reporting Ub sites identified on the basis of a Gly-Gly residue.

Antibodies have been developed for the Lys-epsilon-Gly-Gly (K-ε-GG) residue, and the use of this antibody has been shown to be a viable strategy for enrichment and identification of ubiquitinated peptides. Most recently, the K-ε-GG antibody was used to study the effects of proteasomal and DUB inhibition in Jurkat cells, and 5533 K-ε-GG peptides were identified (Udeshi et al. 2012). Another recent report identified diagnostic b₂' and a₁' fragment ions for K-ε-GG peptides which were labeled with formaldehyde-D2 and NaCNBH₃ (Chicooree et al., 2013). These ions could prove to be valuable in the development of targeted MS-based strategies for the identification of ubiquitinated peptides.

6.6 Lipid Modifications of Proteins

The three major types of lipid modifications of proteins include covalent modifications by fatty acids (N-myristoylation, S- and N-palmytoylation), isoprenoids, and glycosylphosphatidyl inositol (GPI). These modifications may occur separately or in combinations, e.g. myristate + palmitate, palmitate + cholesterol, or farnesyl + palmitate. Some less frequent types of protein lipidation have been also described, such as cholesterol esterification of Hedgehog (Hh) proteins occurring in the ER during autoprocessing (Porter et al. 1996; Ryan and Chiang 2012; Grover et al. 2011; Palm et al. 2013), or direct attachment of a phospholipid phosphatidylethanolamine via an amide bond to the yeast protein Atg8 (Ichimura et al. 2000; Kirisako et al. 2000) and its mammalian homolog LC3 (Kabeya et al. 2004) during autophagy; however, the currently known examples are restricted to small groups of proteins.

The covalent linkage between a protein and either thioester-linked palmitate or a GPI anchor can be broken by the actions of thioesterases: cytosolic acyl protein thioesterase 1 (APT1) (Duncan and Gilman 1998; Zeidman et al. 2009) and lysosomal palmitoyl-protein thioesterase 1 (PPT1) (Camp and Hofmann 1993; Camp et al. 1994), and phospholipases (Low and Prasad 1988; Davitz et al. 1989; Metz et al. 1994), respectively. By contrast, neither myristate nor the isoprenoids farnesyl or geranylgeranyl are physically removed from a modified protein. Instead, some proteins sequester these lipophilic groups within a hydrophobic cleft, effectively shielding them from the aqueous milieu (Zozulya and Stryer 1992).

Until recently detection and analysis of lipid-modified proteins by traditional proteomic approaches has been relatively difficult and was mostly confined to studying individual purified proteins. Most methods relied on metabolic incorporation of radioactively labeled precursors into the proteins with subsequent detection, which required long exposure times, typically 1–3 months. This, as well as relatively high cost, use of hazardous reagents, and relatively low sensitivity made such methods unsuitable for high-throughput analysis of lipid-modified proteins on a proteome scale (Resh 2006; Martin et al. 2008). Application of mass-spectrometry for studying lipid modifications was also focused on individual lipidated proteins

due to their relatively low representation and abundance in the total cellular proteome. The conventional separation techniques used for characterizing the global cellular proteome are not ideally suited for the recovery of highly hydrophobic lipidated peptides (Tom and Martin 2013). This problem may be partially overcome by optimizing the fractionation strategies (Ujihara et al. 2008; Serebryakova et al. 2011; Wotske et al. 2012). However, the major advance in the field came with the introduction of selective chemical labeling of the lipid modification sites.

The first global characterization of protein palmitoylation was performed in yeast (Roth et al. 2006) using acyl biotin exchange (ABE) enrichment technique, which involves alkylation of free thiol groups by N-ethylmaleimide or methyl methanethiosulfonate with subsequent cleavage of thioesters by hydroxylamine (HA) and labeling the exposed cysteine residues with a biotin analogue for further detection or affinity enrichment (Drisdell and Green 2004). This approach combined with multidimensional liquid chromatography (multidimensional protein identification technology (MudPIT) (Washburn et al. 2001) has allowed identification of 47 proteins, including 12 of the 14 previously known. Using deletion mutants for each of the seven yeast DHCC genes and their combinations, the authors found a significant overlap in substrate specificity between the different DHCC isoforms (Roth et al. 2006). This observation was later confirmed by Emmer and co-authors for *Trypanosoma brucei*. Using the same approach, they have identified a total of 124 palmitoylated proteins with an estimated false discovery rate of 1.0 % (Emmer et al. 2011). Kang and co-authors used a combination of ABE enrichment, MudPit and semi-quantitative analysis by Western blotting to study palmitoylated neuronal proteins; they identified the majority of the previously known proteins, as well as a significant number of novel putative protein substrates of palmitoylation (113 candidates in the high confidence group and 318 in the lower confidence group). Palmitoylation has been tested and confirmed for 21 of the newly identified proteins (Kang et al. 2008). A similar strategy was used to explore the protein palmitoylome and analyze DHHC substrate specificity in human endothelial cells (Marin et al. 2012) and macrophages (Merrick et al. 2011). The ABE approach has been also applied to characterize the distribution of palmitoylated proteins between the lipid rafts and non-raft membrane domains in a prostate cancer cell line. The authors applied a combined strategy for palmitoyl protein identification and site characterization (PalmPISC) which involved parallel SDS-PAGE separation and in-gel digestion of individual streptavidin-captured proteins and analysis of biotinylated peptides purified after in-solution digestion of the total protein extract. They have identified 67 known and 331 novel candidate S-acylated proteins, as well as the localization of 25 known and 143 novel candidate S-acylation sites (Yang et al. 2010). The PalmPISC strategy was also used to analyze protein palmitoylation in human platelets (Dowal et al. 2011).

All of these studies relied on spectral counting for relative quantification of proteins present in the experimental (HA-treated) and control (mock-treated) samples to identify proteins specifically enriched by biotin tag capture, with an arbitrary cutoff for the ratio of spectral counts under HA+ versus HA- conditions. Samples were normalized by the spectral counts of co-purifying contaminants present in

both types of samples; however, their content between samples may vary, potentially leading to inaccurate estimates. Zhang and co-authors used a quantitative ABE approach with differential isotope-coded affinity labeling by biotinylated, thiol-reactive isotope-coded affinity tag (ICAT) reagents to identify the protein substrates of the DHHC2 palmitoyltransferase in HeLa cells. The ratios of ABE-enriched proteins from cells with DHHC2 knockdown and control cells, labeled (H) and light (L) ICAT reagents, were calculated using combined peak areas for mutually shared isotopic masses (Zhang et al. 2008).

Quantitative comparison performed with 4-plex isobaric tags for relative and absolute quantitation (iTRAQ) (Hemsley et al. 2013) allowed quantitative comparison between the samples from wild-type *Arabidopsis* and a *tip1-2* mutant deficient for one of the palmitoyltransferases. This method facilitated reliable identification of a total of 561 putative S-acylated proteins and 103 candidate protein targets of the TIP1 palmitoyltransferase.

The acyl-biotin exchange and similar techniques have significantly advanced the proteomic characterization of palmitoylated proteins; they are relatively simple and inexpensive. However, this approach has certain inherent limitations and pitfalls. While it is well suited for studying S-acylated proteins due to the chemical lability of the thioester bond, it cannot be applied for labeling the sites of N-myristoylation, since the lipid moiety is attached to proteins through the stable amide bond. The specificity of palmitoylated protein identification is strongly dependent on the efficiency of the alkylation for complete blockage of all free thiol groups; high numbers of false-positive initial protein identifications which were discarded on the statistical basis could be explained by incomplete alkylation prior to the affinity enrichment step (Emmer et al. 2011; Merrick et al. 2011). Enzymes which use thioester-linked acyl intermediates in their reaction mechanism are another source of false-positive identifications (Yang et al. 2010). An alternative chemical approach to labeling and enrichment of lipid-modified proteins relies on metabolic labeling by bioorthogonal analogs of fatty acids or isoprenoids which contain a reactive group, usually a terminal alkyne or azide, which is used for fluorescent tagging or affinity labeling. This principle was first applied to the analysis of protein farnesylation in COS-1 cells by using an azido-farnesyl analog and subsequently conjugating it through the azide group to biotinylated phosphine capture reagent (bPPCR) using Staudinger ligation reaction. Affinity purification and proteomic analysis of the conjugated proteins led to the identification of 18 farnesylated proteins (Kho et al. 2004). The same chemical approach was used to identify palmitoylated proteins by metabolic incorporation of synthetic azido-tetradecanoic acid, an isosteric analog of palmitate (Kostiuk et al. 2008), as well as for labeling of geranylgeranylated proteins using an azido-geranylgeranyl analog with subsequent fluorescent tagging and detection (Chan et al. 2009).

Another type of bioorthogonal probe, an isosteric analog of palmitate with an ω -terminal acetylene group, 17-octadecynoic acid (17-ODYA), was used for profiling protein palmitoylation in Jurkat T cells, yielding identification of approximately 125 predicted palmitoylated proteins belonging to different functional groups. Like azido-tetradecanoic acid, it was efficiently incorporated by cellular palmitoylation

machinery into the protein substrates, but the fluorescent and affinity tags for detection and enrichment were linked through the Cu(I)-catalyzed azide-alkyne Huisgen cycloaddition reaction (Martin and Cravatt 2009).

Bio-orthogonal approaches allow temporal control of probe incorporation for pulse–chase analysis. Since they do not require thiol reduction and alkylation and multiple precipitation steps, they are more suitable for analyzing smaller samples. Their overall specificity for the targeted lipid modifications is also higher, since it does not depend on thorough modification and shielding of all the free thiol groups in the analyzed proteins; however, false-positive identification of lipid-modified proteins is not totally excluded (Martin 2013). A combined strategy of identification of palmitoylated proteins in *Plasmodium falciparum*, in which the ABE approach and bio-orthogonal metabolic labeling were applied in parallel experiments proved useful for cross-validation of the resulting hits (Jones et al. 2012).

6.7 Inference of PTMs from MS Data

PTMs play an essential role in the protein's destiny and its function. However, it is challenging to identify them in the complex samples, even if the additional enrichment steps are applied. Widely-used search engines such as Mascot (Perkins et al. 1999), X!Tandem (Craig and Beavis 2003), OMSSA (Geer et al. 2004) or SEQUEST (Eng et al. 1994) can routinely identify only a restricted number of user predefined modifications. Standard database search algorithms employ protein amino acid sequences for *in silico* digestion, according to protease cleavage rules, into peptide fragments. Theoretical spectra are matched with empirical spectra and similarities are calculated. Both fixed and variable modifications are specified by the user prior to the search, forcing the search engines to align spectra with the mass shift of the modification (Potthast et al. 2007; Savitski et al. 2011; Bandeira et al. 2007). Typically, oxidation of methionine is specified as a variable modification. However, the inclusion of increased numbers of variable modifications causes search space to expand exponentially. This leads to longer processing times and a greater number of false positive assignments. Thus, platforms developed for the unrestricted identification of modified peptides execute a simple workflow (Ahrne et al. 2010) that can be described as: extraction of candidate peptides/proteins, matching the theoretical spectra and probability assignment (in general, mimicking the standard database search for the peptide identifications). Examples of such software are InsPecT (Tanner et al. 2005), Popitam (Hernandez et al. 2003), P-Mod (Hansen et al. 2005), VEMS 3.0 (Matthiesen et al. 2005), ModifiComb (Savitski et al. 2006), OpenSea (Searle et al. 2005).

The first step of the pipeline is focused on the reduction of the target database to improve matching scores. Most algorithms employ multiple round processing and sequence tag extraction, which can be used separately or in combination. During multiple rounds processing, the data is searched in two rounds to discard peptides which cannot be identified with sufficient confidence. At first, very strict rules are

applied, allowing for one missed cleavage and one or two variable modifications. In round two, the list of significantly identified proteins from the previous step is screened again but with greater tolerance for PTMs. The Bonanza algorithm (Falkner et al. 2008) uses a similar approach, but is applied to search a spectral library. The algorithm assumes that unmodified and modified peptides have the same fragmentation patterns. Sequence tag extraction is based on a *de novo* sequencing algorithm (Dancik et al. 1999; Fernandez-de-Cossio et al. 2000; Johnson and Taylor 2002) to identify a three-four amino acid long peptide “tag” (Mann and Wilm 1994). InsPecT (Tanner et al. 2005) implements such algorithm followed by a trie-based scan of a database. Another way to improve the filtering process is to split the spectrum into intervals and extract the candidate peptides sequentially from each of them based on the reference database, so-called SIMS (Liu et al. 2008). The benefit of the database reduction step is that it decreases the search time and it can be performed externally. For instance, Swiss Protein Identification Toolbox, SwissPIT (Quandt et al. 2009), combines multiple identification tools to create a concise protein database which is then transferred to a PTM search engine such as Popitam or InsPecT.

In the matching step, a theoretical spectrum of *b*- and *y*-ions is generated for each candidate peptide. Usually, the counts for shared peaks between compared spectra are their similarity measure (Craig and Beavis 2003; Geer et al. 2004). Spectral libraries, on the other hand, contain experimental data which include information on the peak intensities. The use of spectral libraries therefore provides better scoring discriminants. QuickMod was designed as a tool for modification spectral library search (Ahrne et al. 2011). However, most algorithms function on a simple assumption of the similarity of the fragmentation patterns, which is not valid in some cases. The mass of the glutamic acid is the same as the mass of the methylated aspartic acid, which would generate the same theoretical spectrum. For this reason, some tools refer to the modification databases such as Unimod (Creasy and Cottrell 2004), DeltaMass (<http://www.abrf.org/index.cfm/dm.home>) or RESID (Garavelli 2004). Some of the modifications cause neutral losses during fragmentation or produce a diagnostic ion which could be considered during the PTM assignment and could help to distinguish between such modifications as lysine acetylation and lysine trimethylation. This algorithm is implemented in VEMS 3.0.

The last step of the standard PTM search algorithm is designed to refine the results. The approach proposed by Tsur et al. (2005) is based on the reasonable assumption that erroneous modification assignments would be distributed randomly throughout the dataset when scanning through all possible mass shifts for each amino acid. Therefore, only peptides with the modifications assigned to an amino acid reported multiple times are considered true positives. In the simplest case, a peptide contains only one possible site of modification but often there are several that could be modified. For instance, methionine residues are less common amino acids and are usually present only once in the peptide, making methionine oxidation site assignment unnecessary. In contrast, when anticipating for proline and tryptophan oxidation as well, the number of potential sites increases dramatically, making the differentiation between these residues more important. While the standard

database search yields reasonable results in defining whether a peptide is modified or not, confident modification site localization may not be delivered. Currently, there are a number of commercial and in-house designed algorithms addressing this issue. Some of these tools are fully integrated into available MS/MS search engines, whereas others are independent isolated platforms performing only the modification site localization.

Generally, all the tools can be divided in two major groups: the ones that estimate the chance of a given peak to be matched randomly or calculate the score reflecting the difference between peptide identifications with various site localizations. The most known and probably the earliest tool in this field was A-Score (Beausoleil et al. 2006) for the SEQUEST search engine, originally designed to handle low resolution data. This algorithm calculates the probability of a site assignment based on the number of unique peaks for distinguishing between two possible sites *b*- and *y*-ions within a 100 Da window, which is then \log_{10} transformed and multiplied by -10 . A similar approach is implemented in InsPecT, SloMo (Albuquerque et al. 2008), and Phosphinator (Phanstiel et al. 2011). PTM Score, developed for the Andromeda search engine, uses the same algorithm but the scores are calculated for each potential site of the peptide and assuming that peptide is modified, the total probability of the modification and therefore all the scores is 100 % (Olsen et al. 2006). A bottleneck in these methods is the erroneous assumption that the chance to match a given mass difference is equally random. However, there are masses that match various amino acids with different modification, and there are masses that can't be matched to any combination. The PhosphoRS scoring algorithm addresses this issue by enabling the extraction of different numbers of peaks, from different spectral areas, within a defined (100 Da) *m/z* window, using the user-defined mass tolerance (Taus et al. 2011). Therefore, this tool is appropriate for the analysis of data obtained in a high mass accuracy instrument.

Other types of scoring algorithms are Mascot Delta Score (Savitski et al. 2011), the SLIP score within Protein Prospector (Baker et al. 2011), and variable modification localization scoring in Spectrum Mill (Agilent). These algorithms calculate the site localization reliability based on the difference between protein identification probabilities. In standard database searches, all potential modification sites are considered and the probability of the correct peptide assignment is estimated. The \log_{10} transformed difference between those values defines the score for each site. However, Spectrum Mill uses the number of matched peaks, their type and the relative intensity of unmatched peaks to estimate the score. The major difference between the three methods is the number of extracted peaks. The SLIP score and Spectrum Mill use fixed number of peaks per spectrum (40 and 25 respectively) whereas the number of peaks extracted with Mascot Delta Score varies to keep peptide identification confidence at the optimal level. Additionally, both the SLIP scoring tool and Spectrum Mill are implemented within the corresponding search engines, enabling routine analysis of site modification localization.

However, the results still require verification because existing tools are not 100 % efficient. The amount of data produced with the current technology in shotgun proteomics is too large to be manually validated, pushing the development of tools

and algorithms to estimate, analogous to the false discovery rate (FDR) for peptide identification measures, false localization rate (FLR). FDR measures are calculated based on the ratio between the number of false positives identified in a decoy database to the total number of identified peptides. It is assumed that their matching frequencies are the same. Unfortunately, it is nearly impossible to have similar estimates for the FLR, because the localization sites have to be known *a priori* to have an exact measure. Thus, there is insufficient information to calculate FLR for all the potential modification sites in a specific dataset. Additionally, identification of the peptide with an erroneously assigned site is not a random match; in contrast, it is a similar event to the correct assignment, which makes the use of decoy sequences an inefficient approach. The latest advance in this field is the Batch-Tag search engine that implements FLR estimation using amino acids that biologically could not be modified (Baker et al. 2010). During the decoy search, glutamate and proline residues were allowed to be phosphorylated. Proline phosphorylation does not occur in nature and phosphorylated glutamate exists only for a very short period of time as intermediate during the biosynthesis of glutamate and proline, making it unlikely to be detected in proteomic studies. In this way, all modifications assigned to proline or glutamate will be incorrect, enabling the estimation of the FLR. There is no direct correlation between a high peptide match score and FLR, because peptides with high scores could still be found to have high FLRs. Also, there is a higher risk of incorrect site assignment when there are two potential sites locating close to each other.

While in the case of phosphorylation the situation is more or less clear, in order to fully understand the signaling mechanisms and cellular responses comprehensive analysis of the other modification types is essential (Wang et al. 2007). Even though suitability claims for general PTM assignments for the previously described algorithms have been made, they are best suited for the phosphopeptide analysis.

In the case of ubiquitination, more advanced methods are required to discern between ubiquitin and ubiquitin-like peptide modifiers. Because the modification (Ub/Ubl) is a protein by nature, it is digested and fragmented during the MS/MS analysis, making spectral interpretation difficult. Spectra produced by Ub-modified peptides include *b*- and *y*-ions of the target peptide and *b*- and *y*-ions of the Ub/Ubl itself. SUMOylation pattern recognition tools may be used to identify peptide modifiers, as in SUMmOn, which considers only the most intense peaks within a 100 Da window (Pedrioli et al. 2006). The algorithm by Kang *et al.* is applicable for unrestricted PTM identification as well as Ub/Ubl modifier (Kang and Yi 2011). The algorithm consists of four stages of PTM identification and two stages of peak matching of Ub/Ubl *b*-ions with the measured peaks, and matching Ub/Ubl *y*-ions with mass shift classes. The differences between all measured peaks and theoretical fragment ions are calculated and divided into mass shift classes which are then filtered based on their intensity, mass deviation, and the number of mass differences in the class. Usually, Ub/Ubl identification relies on Gly-Gly or Leu-Arg-Gly-Gly mass shift on Lys (but not Ub *y*-ions such as Gly and Arg-Gly-Gly) and *b*-ions of free (not attached to target peptide) Ub/Ubls are mainly used for that. However, cysteines alkylated with iodoacetamide during sample preparation have the same

mass as Gly-Gly, causing erroneous identification of Ub/Ubls (Jeram et al. 2009; Witze et al. 2007). In order to generate the sequence of attached Ub/Ubls the algorithm builds multiple mass shift paths based on matched mass shift classes and mass shifts of theoretical Ub/Ubl y -ions. All known and putative Ub/Ubl proteins used to evaluate the program were identified with 91 % accuracy when anticipating only for 1 PTM and 53 % for 2 PTMs.

Protein modification by glycosylation is important to consider. However, the complexity of the structure is significantly higher due to the branched nature of glycans, with different linkages and site isomers/isobars which differ only in their stereochemistry. The presence of such complex structures in the samples complicates the data search dramatically. There are few strategies to address this issue. Library-based sequencing tools (GlycosidIQ and SimGlycan) (Joshi et al. 2004; Apte and Meitei 2010), generate theoretical spectra for each glycan structure in the library and then match them to the measured spectrum, providing a score. Several approaches have emerged to process MSⁿ tandem mass spectrometry data. The saccharide topology analysis tool STAT (Gaucher et al. 2000) compares the list of all plausible oligosaccharide moieties for a predefined m/z , charge and product ion mass with the experimental spectrum, and provides the evaluation of the match. Similarly, Oscar (Lapadula et al. 2005), StrOligo (Ethier et al. 2003), and GlycoFragment (Lohmann and von der Lieth 2003), generate candidate structures from the predefined precursor ion and estimated composition but apply biosynthetic rule restrictions. Another way to perform the search is to match the spectra against a spectral library of oligosaccharides (Kameyama et al. 2005; Zhang et al. 2005). A *de novo*-based sequencing tool, GLYCH (Tang et al. 2005), allows the tree structure of a number of monosaccharide residues, maximizing the number of theoretical ions. Various structural solutions are then evaluated and ranked, taking in consideration one and two stages of fragmentation.

There is a demand for high quality empirical databases as well as technical infrastructure for glycomics. GlycomeDB (Ranzinger et al. 2011), GlycoSuiteDB (Cooper et al. 2001b; 2003), EUROCarbDB (von der Lieth et al. 2011), SWEET-DB (Loss et al. 2002), BOLD (Cooper et al. 1999), and KEGG (Hashimoto et al. 2006) are widely known and often used for glycan searches. GlycoSuiteDB is now included in UniCarbKB (<http://www.unicarbkb.org>), GlycoWorkbench was designed by EUROCardDB initiative to evaluate manual spectrum annotation using the same approach as described above (Ceroni et al. 2008). For easier and faster structural assembly, it contains an intuitive visual editor, GlycanBuilder (Ceroni et al. 2007).

A completely different method to analyze glycans, a combinatorial approach, may be employed in glycan studies. GlycoMod (Cooper et al. 2001a) and Glycopeakfinder (Goldberg et al. 2005) allow *de novo* assignment of glycan composition from a single mass measurement. No prior knowledge of the biological background or fragmentation technique required. However, the number of possible compositions matching certain mass increases exponentially with the number of allowed monomers, leading to the development of tools that consider taxonomic and glyco-biological background, such as Cartoonist (Goldberg et al. 2005) and Retrosynthetic Glycan Network Libraries (Kronewitter et al. 2009). Cartoonist, designed to analyze

MALDI-MS data, generates all plausible mammalian-synthesized N-linked glycan topologies using a manually compiled library of archetypes.

Despite the large variety of tools developed for glycoanalysis, there is little focus on raw spectral processing. The Glycolyzer (Kronewitter et al. 2012), designed on the basis of SysBioWare (Vakhrushev et al. 2009), is an integrated annotation program for glycan biomarker discovery. It contains all the basic components (background subtraction, peak detection, noise removal and data processing) as well as calibration, theoretical retrosynthetic library based glycan annotation and statistical hypothesis testing. The workflow uses FT-ICR MS data for the input.

The previously described bioinformatic tools are suitable for the oligosaccharide structural assignments. However, Byonic (Bern et al. 2012) and GlycoPeptideSearch (Pompach et al. 2012) combine the known glycan analysis methods and the proteomics search or user-specified potential glycosylated peptides. GlycoSearch (Kletter et al. 2013) is a related tool that is used for glycan binding motif analysis of lectins.

Finally, we will describe the development of bioinformatics tools for the analysis of lipid modifications of proteins. Sites of modification may be predicted *in silico* using numerous amino acid sequence-based tools for various lipids (partly available via <http://mendel.imp.ac.at/>). For proteomic purposes, lipoproteins are isolated from the sample and enriched narrowing the scope of the work for a standard peptide identification. However, for more information on structure, function, biosynthesis and association with certain protein (pathway) one could refer to LIPID Maps (<http://www.lipidmaps.org/>).

In spite of the rapid development of bioinformatics for automated identification and site localization of modifications, verification is still mostly manual. For that purpose, knowledge-based libraries of all modifications such as DeltaMass and UniMod are very useful, as well as PhosphoSitePlus (<http://www.phosphosite.org/>), Phospho.ELM (<http://phospho.elm.eu.org/>), PHOSIDA (Gnad et al. 2011) and METLIN (a metabolite database; <http://metlin.scripps.edu>) (Smith et al. 2005). However, in order to be able to identify novel modifications, *de novo* sequencing algorithms implemented in PepNovo (Frank and Pevzner 2005) or PEAKS (Ma et al. 2003) would be quite useful in this respect.

6.8 Summary

The large number of types PTMs that have been identified create an enormous challenge to proteomic studies modifications. The challenges are related to the chemical nature of PTMs, their microheterogeneity, and site localization. In this chapter, we highlighted the biological importance of the most common types of PTMS, namely, protein acetylation, phosphorylation, glycosylation, ubiquitination, and lipidation. We also provided an overview of separation methods, mass spectrometric analysis, and recent developments in bioinformatic strategies to analyze PTMs on a proteomic scale.

References

- Ahrne E, Muller M, Lisacek F. Unrestricted identification of modified proteins using MS/MS. *Proteomics*. 2010;10:671–86.
- Ahrne E, Nikitin F, Lisacek F, Muller M. QuickMod: a tool for open modification spectrum library searches. *J Proteome Res*. 2011;10:2913–21.
- Albuquerque CP, Smolka MB, Payne SH, Bafna V, Eng J, Zhou H. A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol Cell Proteomics*. 2008;7:1389–96.
- Alley WR, Mechref Y, Novotny MV. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Commun Mass Spectrom*. 2009;23:161–70.
- Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J Chromatogr*. 1990;9:177–96.
- Alpert AJ. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal Chem*. 2007;80:62–76.
- Alpert AJ. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal Chem*. 2008;80:62–76.
- An HJ, Peavy TR, Hedrick JL, Lebrilla CB. Determination of N-glycosylation sites and site heterogeneity in glycoproteins. *Anal Chem*. 2003;75:5628–37.
- Apte A, Meitei NS. Bioinformatics in glycomics: glycan characterization with mass spectrometric data using SimGlycan. *Methods Mol Biol*. 2010;600:269–81.
- Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophys Acta (BBA) Gen Subj*. 1999;1473:4–8.
- Atwood JA, Minning T, Ludolf F, Nuccio A, Weatherly DB, Alvarez-Manilla G, Tarleton R, Orlando R. Glycoproteomics of *Trypanosoma cruzi* trypomastigotes using subcellular fractionation, lectin affinity, and stable isotope labeling. *J Proteome Res*. 2006;5:3376–84.
- Bachmair A, Finley D, Varshavsky A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science*. 1986;234:179–86.
- Baker PR, Medzihradsky KF, Chalkley RJ. Improving software performance for peptide electron transfer dissociation data analysis by implementation of charge state- and sequence-dependent scoring. *Mol Cell Proteomics*. 2010;9:1795–803.
- Baker PR, Trinidad JC, Chalkley RJ. Modification site localization scoring integrated into a search engine. *Mol Cell Proteomics*. 2011;10.
- Bandeira N, Tsur D, Frank A, Pevzner P. Protein identification by spectral networks analysis. *Proc Natl Acad Sci U S A*. 2007;104:6140–5.
- Bannister AJ, Miska EA, Görllich D, Kouzarides T. Acetylation of importin- α nuclear import factors by CBP/p300. *Curr Biol*. 2000;10:467–70.
- Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol*. 2006;24:1285–92.
- Bern M, Kil YJ, Ecker C. Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics*. 2012. Chapter 13, Unit 13 20.
- Boersema P, Mohammed S, Heck AR. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Anal Bioanal Chem*. 2008;391:151–9.
- Bunkenborg J, Pilch BJ, Podtelejnikov AV, Wisniewski JR. Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics*. 2004;4:454–65.
- Camp LA, Hofmann SL. Purification and properties of a palmitoyl-protein thioesterase that cleaves palmitate from H-Ras. *J Biol Chem*. 1993;268:22566–74.
- Camp LA, Verkruyse LA, Afendis SJ, Slaughter CA, Hofmann SL. Molecular cloning and expression of palmitoyl-protein thioesterase. *J Biol Chem*. 1994;269:23212–19.
- Černý M, Skalák J, Cerna H, Brzobohatý B. Advances in purification and separation of posttranslationally modified proteins. *J Proteomics*. 2013;92:2–27.

- Ceroni A, Dell A, Haslam SM. The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. *Sour Code Biol Med*. 2007;2:3.
- Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res*. 2008;7:1650–9.
- Chan LN, Hart C, Guo L, Nyberg T, Davies BS, Fong LG, Young SG, Agnew BJ, Tamanoi F. A novel approach to tag and identify geranylgeranylated proteins. *Electrophoresis*. 2009;30:3598–606.
- Chicooree N, Connolly Y, Tan CT, Malliri A, Li Y, Smith DL, Griffiths JR. Enhanced detection of ubiquitin isopeptides using reductive methylation. *J Am Soc Mass Spectrom*. 2013;24(3):421–30.
- Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
- Cooper CA, Wilkins MR, Williams KL, Packer NH. BOLD—a biological O-linked glycan database. *Electrophoresis*. 1999;20:3589–98.
- Cooper CA, Gasteiger E, Packer NH. GlycoMod—a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics*. 2001a;1:340–9.
- Cooper CA, Harrison MJ, Wilkins MR, Packer NH. GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res*. 2001b;29:332–5.
- Cooper CA, Joshi HJ, Harrison MJ, Wilkins MR, Packer NH. GlycoSuiteDB: a curated relational database of glycoprotein glycan structures and their biological sources. 2003 update. *Nucleic Acids Res*. 2003;31:511–13.
- Craig R, Beavis RC. A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Commun Mass Spectrom*. 2003;17:2310–16.
- Creasy DM, Cottrell JS. Unimod: protein modifications for mass spectrometry. *Proteomics*. 2004;4:1534–6.
- Crimmins DL, Gorka J, Thoma RS, Schwartz BD. Peptide characterization with a sulfoethyl aspartamide column. *J Chromatogr*. 1988;443:63–71.
- Dancik V, Addona TA, Clauser KR, Vath JE, Pevzner PA. De novo peptide sequencing via tandem mass spectrometry. *J Comput Biol*. 1999;6:327–42.
- Davies M, Smith KD, Harbin AM, Hounsell EF. High-performance liquid chromatography of oligosaccharide alditols and glycopeptides on a graphitized carbon column. *J Chromatogr A*. 1992;609:125–31.
- Davitz MA, Hom J, Schenkman S. Purification of a glycosyl-phosphatidylinositol-specific phospholipase D from human plasma. *J Biol Chem*. 1989;264:13760–4.
- Dell A, Morris HR. Glycoprotein structure determination by mass spectrometry. *Science*. 2001;291:2351–6.
- Dephoure N, Gould KL, Gygi SP, Kellogg DR. Mapping and analysis of phosphorylation sites: a quick guide for cell biologists. *Mol Biol Cell*. 2013;24:535–42.
- Di Palma S, Hennrich ML, Heck AJR, Mohammed S. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J Proteomics*. 2012;75:3791–813.
- Ding W, Hill JJ, Kelly J. Selective enrichment of glycopeptides from glycoprotein digests using ion-pairing normal-phase liquid chromatography. *Anal Chem*. 2007;79:8891–9.
- Dormeyer W, Mohammed S, Breukelen BV, Krijgsveld J, Heck AJR. Targeted analysis of protein termini. *J Proteome Res*. 2007;6:4634–45.
- Dowal L, Yang W, Freeman MR, Steen H, Flaumenhaft R. Proteomic analysis of palmitoylated platelet proteins. *Blood*. 2011;118:e62–73.
- Drisdel RC, Green WN. Labeling and quantifying sites of protein palmitoylation. *Biotechniques*. 2004;36:276–85.
- Dube DH, Bertozzi CR. Glycans in cancer and inflammation [mdash] potential for therapeutics and diagnostics. *Nat Rev Drug Discov*. 2005;4:477–88.
- Duncan JA, Gilman AG. A cytoplasmic acyl-protein thioesterase that removes palmitate from G protein alpha subunits and p21(RAS). *J Biol Chem*. 1998;273:15830–7.
- Eberharter A, Becker PB. Histone acetylation: a switch between repressive and permissive chromatin. *EMBO Rep*. 2002;3:224–9.

- Edwards AVG, Edwards GJ, Schwämmle V, Saxtorph H, Larsen MR. Spatial and temporal effects in protein post-translational modification distributions in the developing mouse brain. *J Proteome Res.* 2014;13(1):260–7.
- Emmer BT, Nakayasu ES, Souther C, Choi H, Sobreira TJ, Epting CL, Nesvizhskii AI, Almeida IC, Engman DM. Global analysis of protein palmitoylation in African trypanosomes. *Eukaryot Cell.* 2011;10:455–63.
- Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J Am Soc Mass Spectrom.* 1994;5:976–89.
- Ethier M, Saba JA, Spearman M, Krokhin O, Butler M, Ens W, Standing KG, Perrault H. Application of the StrOligo algorithm for the automated structure assignment of complex N-linked glycans from glycoproteins using tandem mass spectrometry. *Rapid Commun Mass Spectrom.* 2003;17:2713–20.
- Eyrich B, Sickmann A, Zahedi RP. Catch me if you can: mass spectrometry-based phosphoproteomics and quantification strategies. *Proteomics.* 2011;11:554–70.
- Falkner JA, Falkner JW, Yocum AK, Andrews PC. A spectral clustering approach to MS/MS identification of post-translational modifications. *J Proteome Res.* 2008;7:4614–22.
- Fan JQ, Kondo A, Kato I, Lee YC. High-performance liquid chromatography of glycopeptides and oligosaccharides on graphitized carbon columns. *Anal Biochem.* 1994;219:224–9.
- Ferluga S, Hantgan R, Goldgur Y, Himanen JP, Nikolov DB, Debinski W. Biological and structural characterization of glycosylation on ephrin-A1, a preferred ligand for EphA2 receptor tyrosine kinase. *J Biol Chem.* 2013;288:18448–57.
- Fernandez-De-Cossio J, Gonzalez J, Satomi Y, Shima T, Okumura N, Besada V, Betancourt L, Padron G, Shimonishi Y, Takao T. Automated interpretation of low-energy collision-induced dissociation spectra by SeqMS, a software aid for de novo sequencing by tandem mass spectrometry. *Electrophoresis.* 2000;21:1694–9.
- Ficarro SB, Adelmant GO, Tomar MN, Zhang Y, Cheng VJ, Marto JA. Magnetic bead processor for rapid evaluation and optimization of parameters for phosphopeptide enrichment. *Anal Chem.* 2009;81:4566–75.
- Frank A, Pevzner P. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem.* 2005;77:964–73.
- Garavelli JS. The RESID database of protein modifications as a resource and annotation tool. *Proteomics.* 2004;4:1527–33.
- Gates M, Tomer K, Detering L. Comparison of metal and metal oxide media for phosphopeptide enrichment prior to mass spectrometric analyses. *J Am Soc Mass Spectrom.* 2010;21:1649–59.
- Gaucher SP, Cancilla MT, Phillips NJ, Gibson BW, Leary JA. Mass spectral characterization of lipooligosaccharides from *Haemophilus influenzae* 2019. *Biochemistry.* 2000;39:12406–14.
- Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH. Open mass spectrometry search algorithm. *J Proteome Res.* 2004;3:958–64.
- Gilar M, Olivova P, Daly AE, Gebler JC. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J Sep Sci.* 2005;28:1694–703.
- Gilar M, Yu Y-Q, Ahn J, Fournier J, Gebler JC. Mixed-mode chromatography for fractionation of peptides, phosphopeptides, and sialylated glycopeptides. *J Chromatogr A.* 2008;1191:162–70.
- Glotzer M, Murray AW, Kirschner MW. Cyclin is degraded by the ubiquitin pathway. *Nature.* 1991;349:132–8.
- Glozak MA, Sengupta N, Zhang X, Seto E. Acetylation and deacetylation of non-histone proteins. *Gene.* 2005;363:15–23.
- Gnad F, Gunawardena J, Mann M. PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res.* 2011;39:D253–60.
- Goldberg D, Sutton-Smith M, Paulson J, Dell A. Automatic annotation of matrix-assisted laser desorption/ionization N-glycan spectra. *Proteomics.* 2005;5:865–75.
- Grover VK, Valadez JG, Bowman AB, Cooper MK. Lipid modifications of Sonic hedgehog ligand dictate cellular reception and signal response. *PLoS One.* 2011;6:e21353.
- Gupta S, Seth A, Davis RJ. Transactivation of gene expression by Myc is inhibited by mutation at the phosphorylation sites Thr-58 and Ser-62. *Proc Natl Acad Sci.* 1993;90:3216–20.

- Hägglund P, Bunkenborg J, Elortza F, Jensen ON, 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.* 2004;3:556–66.
- Hägglund P, Matthiesen R, Elortza F, Højrup P, Roepstorff P, Jensen ON, Bunkenborg J. An enzymatic deglycosylation scheme enabling identification of core fucosylated N-glycans and O-glycosylation site mapping of human plasma proteins. *J Proteome Res.* 2007;6:3021–31.
- Hakansson K, Cooper HJ, Emmett MR, Costello CE, Marshall AG, Nilsson CL. Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptide to yield complementary sequence information. *Anal Chem.* 2001;73:4530–6.
- Hansen BT, Davey SW, Ham AJL, Liebler DC. P-mod: an algorithm and software to map modifications to peptide sequences using tandem MS data. *J Proteome Res.* 2005;4:358–68.
- Hao P, Guo T, Sze SK. Simultaneous analysis of proteome, phospho- and glycoproteome of rat kidney tissue with electrostatic repulsion hydrophilic interaction chromatography. *PLoS One.* 2011;6:e16884.
- Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu Rev Biochem.* 2011;80:825–58.
- Hashimoto K, Goto S, Kawano S, Aoki-Kinoshita KF, Ueda N, Hamajima M, Kawasaki T, Kanehisa M. KEGG as a glycome informatics resource. *Glycobiology.* 2006;5:63R–70.
- Heidinger-Pauli JM, Ünal E, Koshland D. Distinct targets of the EcoI acetyltransferase modulate cohesion in S phase and in response to DNA damage. *Mol Cell.* 2009;34:311–21.
- Hemsley PA, Weimar T, Lilley KS, Dupree P, Grierson CS. A proteomic approach identifies many novel palmitoylated proteins in Arabidopsis. *New Phytol.* 2013;197:805–14.
- Hernandez P, Gras R, Frey J, Appel RD. Popitam: towards new heuristic strategies to improve protein identification from tandem mass spectrometry data. *Proteomics.* 2003;3:870–8.
- Himanen JP, Goldgur Y, Miao H, Myshkin E, Guo H, Buck M, Nguyen M, Rajashankar KR, Wang B, Nikolov DB. Ligand recognition by A-class Eph receptors: crystal structures of the EphA2 ligand-binding domain and the EphA2/ephrin-A1 complex. *EMBO Rep.* 2009;10:722–8.
- Himanen JP, Yermekbayeva L, Janes PW, Walker JR, Xu K, Atapattu L, Rajashankar KR, Mensinga A, Lackmann M, Nikolov DB, Dhe-Paganon S. Architecture of Eph receptor clusters. *Proc Natl Acad Sci.* 2010;107:10860–5.
- Horn DM, Ge Y, McLafferty FW. Activated ion electron capture dissociation for mass spectral sequencing of larger (42 KDa) proteins. *Anal Chem.* 2000;72:4778–84.
- Hua S, Hu CY, Kim BJ, Totten SM, Oh MJ, Yun N, Nwosu CC, Yoo JS, Lebrilla CB, An HJ. Glyco-Analytical Multispecific Proteolysis (Glyco-AMP): a simple method for detailed and quantitative glycoproteomic characterization. *J Proteome Res.* 2013;12:4414–23.
- Hwang C-S, Shemorry A, Varshavsky A. N-terminal acetylation of cellular proteins creates specific degradation signals. *Science.* 2010;327:973–7.
- Ichimura Y, Kirisako T, Takao T, Satomi Y, Shimonishi Y, Ishihara N, Mizushima N, Tanida I, Kominami E, Ohsumi M, Noda T, Ohsumi Y. A ubiquitin-like system mediates protein lipidation. *Nature.* 2000;408:488–92.
- Imperiali B, Rickert KW. Conformational implications of asparagine-linked glycosylation. *Proc Natl Acad Sci.* 1995;92:97–101.
- Jadhav T, Wooten MW. Defining an embedded code for protein ubiquitination. *J Proteomics Bioinform.* 2009;2:316.
- Jeram SM, Srikumar T, Pedrioli PG, Raught B. Using mass spectrometry to identify ubiquitin and ubiquitin-like protein conjugation sites. *Proteomics.* 2009;9:922–34.
- Johnson RS, Taylor JA. Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry. *Mol Biotechnol.* 2002;22:301–15.
- Jones ML, Collins MO, Goulding D, Choudhary JS, Rayner JC. Analysis of protein palmitoylation reveals a pervasive role in Plasmodium development and pathogenesis. *Cell Host Microbe.* 2012;12:246–58.
- Joshi HJ, Harrison MJ, Schulz BL, Cooper CA, Packer NH, Karlsson NG. Development of a mass fingerprinting tool for automated interpretation of oligosaccharide fragmentation data. *Proteomics.* 2004;4:1650–64.

- Kabeya Y, Mizushima N, Yamamoto A, Oshitani-Okamoto S, Ohsumi Y, Yoshimori T. LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *J Cell Sci.* 2004;117:2805–12.
- Kameyama A, Kikuchi N, Nakaya S, Ito H, Sato T, Shikanai T, Takahashi Y, Takahashi K, Narimatsu H. A strategy for identification of oligosaccharide structures using observational multistage mass spectral library. *Anal Chem.* 2005;77:4719–25.
- Kang C, Yi GS. Identification of ubiquitin/ubiquitin-like protein modification from tandem mass spectra with various PTMs. *BMC Bioinform.* 2011;12 Suppl 14:S8.
- Kang R, Wan J, Arstikaitis P, Takahashi H, Huang K, Bailey AO, Thompson JX, Roth AF, Drisdell RC, Mastro R, Green WN, Yates JR, Davis 3rd NG, El-Husseini A. Neural palmitoyl-proteomics reveals dynamic synaptic palmitoylation. *Nature.* 2008;456:904–9.
- Kelleher NL, Zubarev RA, Bush K, Furie B, Furie BC, McLafferty FW, Walsh CT. Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. *Anal Chem.* 1999;71:4250–3.
- Kho Y, Kim SC, Jiang C, Barma D, Kwon SW, Cheng J, Jaunbergs J, Weinbaum C, Tamanoi F, Falck J, Zhao Y. A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *Proc Natl Acad Sci U S A.* 2004;101:12479–84.
- Kimura A, Matsubara K, Horikoshi M. A decade of histone acetylation: marking eukaryotic chromosomes with specific codes. *J Biochem.* 2005;138:647–62.
- Kirisako T, Ichimura Y, Okada H, Kabeya Y, Mizushima N, Yoshimori T, Ohsumi M, TAKAO T, Noda T, Ohsumi Y. The reversible modification regulates the membrane-binding state of Apg8/Aut7 essential for autophagy and the cytoplasm to vacuole targeting pathway. *J Cell Biol.* 2000;151:263–76.
- Kletter D, Cao Z, Bern M, Haab B. Determining lectin specificity from glycan array data using motif segregation and GlycoSearch software. *Curr Protoc Chem Biol.* 2013;5:157–69.
- Koizumi K, Okada Y, Fukuda M. High-performance liquid chromatography of mono- and oligosaccharides on a graphitized carbon column. *Carbohydr Res.* 1991;215:67–80.
- Komander D, Clague MJ, Urbe S. Breaking the chains: structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol.* 2009;10:550–63.
- Kostiuk MA, Corvi MM, Keller BO, Plummer G, Prescher JA, Hangauer MJ, Bertozzi CR, Rajaiiah G, Falck JR, Berthiaume LG. Identification of palmitoylated mitochondrial proteins using a bio-orthogonal azido-palmitate analogue. *FASEB J.* 2008;22:721–32.
- Kouzarides T. Acetylation: a regulatory modification to rival phosphorylation? *EMBO J.* 2000;19:1176–9.
- Kronewitter SR, An HJ, de Leoz ML, Lebrilla CB, Miyamoto S, Leiserowitz GS. The development of retrosynthetic glycan libraries to profile and classify the human serum N-linked glycome. *Proteomics.* 2009;9:2986–94.
- Kronewitter SR, DE Leoz ML, Strum JS, An HJ, Dimapasoc LM, Guerrero A, Miyamoto S, Lebrilla CB, Leiserowitz GS. The glycolyzer: automated glycan annotation software for high performance mass spectrometry and its application to ovarian cancer glycan biomarker discovery. *Proteomics.* 2012;12:2523–38.
- Lapadula AJ, Hatcher PJ, Hanneman AJ, Ashline DJ, Zhang H, Reinhold VN. Congruent strategies for carbohydrate sequencing. 3. OSCAR: an algorithm for assigning oligosaccharide topology from MSn data. *Anal Chem.* 2005;77:6271–9.
- Larsen MR, Højrup P, Roepstorff P. Characterization of gel-separated glycoproteins using Two-step proteolytic digestion combined with sequential microcolumns and mass spectrometry. *Mol Cell Proteomics.* 2005;4:107–19.
- Larsen MR, Jensen SS, Jakobsen LA, Heegaard NHH. Exploring the sialome using titanium dioxide chromatography and mass spectrometry. *Mol Cell Proteomics.* 2007;6:1778–87.
- Lewandrowski U, Zahedi RP, Moebius J, Walter U, Sickmann A. Enhanced N-glycosylation site analysis of sialoglycopeptides by strong cation exchange prefractionation applied to platelet plasma membranes. *Mol Cell Proteomics.* 2007;6(11):1933–41.
- Lin R, Zhou X, Huang W, Zhao D, Lu L, Xiong Y, Guan KL, Lei QY. Acetylation control of cancer metabolism. *Curr Pharm Des.* 2014;20(15):2627–33.

- Little DP, Speir JP, Senko MW, O'connor PB, McLafferty FW. Infrared multiphoton dissociation of large multiply-charged ions for biomolecule sequencing. *Anal Chem.* 1994;66:2809–15.
- Liu J, Erassov A, Halina P, Canete M, Nguyen DV, Chung C, Cagney G, Ignatchenko A, Fong V, Emili A. Sequential interval motif search: unrestricted database surveys of global MS/MS data sets for detection of putative post-translational modifications. *Anal Chem.* 2008;80:7846–54.
- Lohmann KK, von der Lieth CW. GLYCO-FRAGMENT: a web tool to support the interpretation of mass spectra of complex carbohydrates. *Proteomics.* 2003;3:2028–35.
- Loss A, Bunsmann P, Bohne A, Schwarzer E, Lang E, von der Lieth CW. SWEET-DB: an attempt to create annotated data collections for carbohydrates. *Nucleic Acids Res.* 2002;30:405–8.
- Low MG, Prasad AR. A phospholipase D specific for the phosphatidylinositol anchor of cell-surface proteins is abundant in plasma. *Proc Natl Acad Sci U S A.* 1988;85:980–4.
- Luo J, Su F, Chen D, Shiloh A, Gu W. Deacetylation of p53 modulates its effect on cell growth and apoptosis. *Nature.* 2000;408:377–81.
- Luo J, Li M, Tang Y, Laszkowska M, Roeder RG, Gu W. Acetylation of p53 augments its site-specific DNA binding both in vitro and in vivo. *Proc Natl Acad Sci U S A.* 2004;101:2259–64.
- Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom.* 2003;17:2337–42.
- Malm D, Nilssen O. Alpha-mannosidosis. *Orphanet J Rare Dis.* 2008;3:21.
- Mann M, Wilm M. Error tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem.* 1994;66:4390–9.
- Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science.* 2002;298:1912–34.
- Manning JM, Popowicz AM, Padovan JC, Chait BT, Manning LR. Intrinsic regulation of hemoglobin expression by variable subunit interface strengths. *FEBS J.* 2012;279:361–9.
- Marin EP, Derakhshan B, Lam TT, Davalos A, Sessa WC. Endothelial cell palmitoylproteomic identifies novel lipid-modified targets and potential substrates for protein acyl transferases. *Circ Res.* 2012;110:1336–44.
- Marshall AG, Hendrickson CL. High-resolution mass spectrometers. In: Yeung SY, Zare RN, editors. *Annual review of analytical chemistry.* Palo Alto: Annual Reviews; 2008.
- Martin BR. Chemical approaches for profiling dynamic palmitoylation. *Biochem Soc Trans.* 2013;41:43–9.
- Martin BR, Cravatt BF. Large-scale profiling of protein palmitoylation in mammalian cells. *Nat Methods.* 2009;6:135–8.
- Martin DD, Vilas GL, Prescher JA, Rajaiah G, Falck JR, Bertozzi CR, Berthiaume LG. Rapid detection, discovery, and identification of post-translationally myristoylated proteins during apoptosis using a bio-orthogonal azidomyristate analog. *FASEB J.* 2008;22:797–806.
- Martin-Rendon E, Blake DJ. Protein glycosylation in disease: new insights into the congenital muscular dystrophies. *Trends Pharmacol Sci.* 2003;24:178–83.
- Matsumoto ML, Wickliffe KE, Dong KC, Yu C, Bosanac I, Bustos D, Phu L, Kirkpatrick DS, Hymowitz SG, Rape M, Kelley RF, Dixit VM. K11-linked polyubiquitination in cell cycle control revealed by a K11 linkage-specific antibody. *Mol Cell.* 2010;39:477–84.
- Matthiesen R, Trelle MB, Hojrup P, Bunkenborg J, Jensen ON. VEMS 3.0: algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. *J Proteome Res.* 2005;4:2338–47.
- Merrick BA, Dhungana S, Williams JG, Aloor JJ, Peddada S, Tomer KB, Fessler MB. Proteomic profiling of S-acylated macrophage proteins identifies a role for palmitoylation in mitochondrial targeting of phospholipid scramblase 3. *Mol Cell Proteomics.* 2011;10:M110.006007.
- Metz CN, Brunner G, Choi-Muira NH, Nguyen H, Gabrilove J, Caras IW, Altszuler N, Rifkin DB, Wilson EL, Davitz MA. Release of GPI-anchored membrane proteins by a cell-associated GPI-specific phospholipase D. *EMBO J.* 1994;13:1741–51.
- Metzger MB, Pruneda JN, Klevit RE, Weissman AM. RING-type E3 ligases: Master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochim Biophys Acta.* 2014;1843(1):47–60.

- Murphy M, Ahn J, Walker KK, Hoffman WH, Evans RM, Levine AJ, George DL. Transcriptional repression by wild-type p53 utilizes histone deacetylases, mediated by interaction with mSin3a. *Genes Dev.* 1999;13:2490–501.
- Mysling S, Palmisano G, Højrup P, Thaysen-Andersen M. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal Chem.* 2010;82:5598–609.
- Nakayasu ES, Ansong C, Brown JN, Yang F, Lopez-Ferrer D, Qian WJ, Smith RD, Adkins JN. Evaluation of selected binding domains for the analysis of ubiquitinated proteomes. *J Am Soc Mass Spectrom.* 2013;24:1214–23.
- Newton K, Matsumoto ML, Wertz IE, Kirkpatrick DS, Lill JR, Tan J, Dugger D, Gordon N, Sidhu SS, Fellouse FA, Komuves L, French DM, Ferrando RE, Lam C, Compaan D, Yu C, Bosanac I, Hymowitz SG, Kelley RF, Dixit VM. Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies. *Cell.* 2008;134:668–78.
- Nilsson CL. Advances in quantitative phosphoproteomics. *Anal Chem.* 2011a;84:735–46.
- Nilsson CL. Lectin techniques for glycoproteomics. *Curr Proteomics.* 2011b;8:248–56.
- Nilsson IM, von Heijne G. Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J Biol Chem.* 1993;268:5798–801.
- Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P, Mann M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell.* 2006;127:635–48.
- Oppermann FS, Gnäd F, Olsen JV, Hornberger R, Greff Z, Keri G, Mann M, Daub H. Large-scale proteomics analysis of the human kinome. *Mol Cell Proteomics.* 2009;8:1751–64.
- Packer NH, Lawson MA, Jardine DR, Redmond JW. A general approach to desalting oligosaccharides released from glycoproteins. *Glycoconj J.* 1998;15:737–47.
- Palm W, Swierczynska MM, Kumari V, Ehrhart-Bornstein M, Bornstein SR, Eaton S. Secretion and signaling activities of lipoprotein-associated hedgehog and non-sterol-modified hedgehog in flies and mammals. *PLoS Biol.* 2013;11:e1001505.
- Palmisano G, Lendal S, Larsen M. Titanium dioxide enrichment of sialic acid-containing glycopeptides. In: Gevaert K, Vandekerckhove J, editors. *Gel-free proteomics*. Totowa: Humana Press; 2011.
- Palmisano G, Melo-Braga MN, Engholm-Keller K, Parker BL, Larsen MR. Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J Proteome Res.* 2012;11:1949–57.
- Pedrioli PG, Raught B, Zhang XD, Rogers R, Aitchison J, Matunis M, Aebersold R. Automated identification of SUMOylation sites using mass spectrometry and SUMOn pattern recognition software. *Nat Methods.* 2006;3:533–9.
- Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, Roelofs J, Finley D, Gygi SP. A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol.* 2003;21:921–6.
- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* 1999;20:3551–67.
- Pfleger CM, Kirschner MW. The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. *Genes Dev.* 2000;14:655–65.
- Phanstiel DH, Brumbaugh J, Wenger CD, Tian SL, Probasco MD, Bailey DJ, Swaney DL, Tervo MA, Bolin JM, Ruotti V, Stewart R, Thomson JA, Coon JJ. Proteomic and phosphoproteomic comparison of human ES and iPSC cells. *Nat Methods.* 2011;8:821–7.
- Pompach P, Chandler KB, Lan R, Edwards N, Goldman R. Semi-automated identification of N-Glycopeptides by hydrophilic interaction chromatography, nano-reverse-phase LC-MS/MS, and glycan database search. *J Proteome Res.* 2012;11:1728–40.
- Porter JA, Young KE, Beachy PA. Cholesterol modification of hedgehog signaling proteins in animal development. *Science.* 1996;274:255–9.
- Potthast F, Gerrits B, Hakkinen J, Rutishauser D, Ahrens CH, Roschitzki B, Baerenfaller K, Munton RP, Walther P, Gehrig P, Seif P, Seeberger PH, Schlapbach R. The mass distance fingerprint: a statistical framework for de novo detection of predominant modifications using high-accuracy mass spectrometry. *J Chromatogr B-Anal Technol Biomed Life Sci.* 2007;854:173–82.

- Quandt A, Masselot A, Hernandez P, Hernandez C, Maffioletti S, Appel RD, Lisacek F. SwissPIT: an workflow-based platform for analyzing tandem-MS spectra using the grid. *Proteomics*. 2009;9:2648–55.
- Ranzinger R, Herget S, von der Lieth CW, Frank M. GlycomeDB—a unified database for carbohydrate structures. *Nucleic Acids Res*. 2011;39:D373–6.
- Reich NC. STAT3 Revs up the powerhouse. *Sci Signal*. 2009;2:pe61.
- Reinhold V, Zhang H, Hanneman A, Ashline D. Toward a platform for comprehensive glycan sequencing. *Mol Cell Proteomics*. 2013;12:866–73.
- Resh MD. Use of analogs and inhibitors to study the functional significance of protein palmitoylation. *Methods*. 2006;40:191–7.
- Rivera CM, Ren B. Mapping human epigenomes. *Cell*. 2013;155:39–55.
- Robinson NE, Robinson A. Deamidation of human proteins. *Proc Natl Acad Sci*. 2001;98:12409–13.
- Rogers S, Wells R, Rechsteiner M. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*. 1986;234:364–8.
- Roth AF, Wan J, Bailey AO, Sun B, Kuchar JA, Green WN, Phinney BS, Yates 3rd JR, Davis NG. Global analysis of protein palmitoylation in yeast. *Cell*. 2006;125:1003–13.
- Rutishauser U. Polysialic acid in the plasticity of the developing and adult vertebrate nervous system. *Nat Rev Neurosci*. 2008;9:26–35.
- Ryan KE, Chiang C. Hedgehog secretion and signal transduction in vertebrates. *J Biol Chem*. 2012;287:17905–13.
- Satomi Y, Shimonishi Y, Takao T. N-glycosylation at Asn491 in the Asn-Xaa-Cys motif of human transferrin. *FEBS Lett*. 2004;576:51–6.
- Savitski MM, Nielsen ML, Zubarev RA. ModifiComb, a new proteomic tool for mapping stoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol Cell Proteomics*. 2006;5:935–48.
- Savitski MM, Lemeer S, Boesche M, Lang M, Mathieson T, Bantscheff M, Kuster B. Confident phosphorylation site localization using the mascot delta score. *Mol Cell Proteomics*. 2011;10:M110.003830.
- Schachter H, Freeze HH. Glycosylation diseases: Quo vadis? *Biochimica et Biophys Acta (BBA) Mol Basis Dis*. 2009;1792:925–30.
- Scheffner M, Kumar S. Mammalian HECT ubiquitin-protein ligases: biological and pathophysiological aspects. *Biochim Biophys Acta*. 2014;1843(1):61–74.
- Schulman BA, Harper JW. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol*. 2009;10:319–31.
- Scott NE, Parker BL, Connolly AM, Paulech J, Edwards AVG, Crossett B, Falconer L, Kolarich D, Djordjevic SP, Højrup P, Packer NH, Larsen MR, Cordwell SJ. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Mol Cell Proteomics*. 2011;10.
- Searle BC, Dasari S, Wilmarth PA, Turner M, Reddy AP, David LL, Nagalla SR. Identification of protein modifications using MS/MS de novo sequencing and the OpenSea alignment algorithm. *J Proteome Res*. 2005;4:546–54.
- Serebryakova MV, Demina IA, Galyamina MA, Kondratov IG, Ladygina VG, Govorun VM. The acylation state of surface lipoproteins of mollicute *Acholeplasma laidlawii*. *J Biol Chem*. 2011;286:22769–76.
- Singh C, Zampronio CG, Creese AJ, Cooper HJ. Higher Energy Collision Dissociation (HCD) Product Ion-Triggered Electron Transfer Dissociation (ETD) mass spectrometry for the analysis of N-linked glycoproteins. *J Proteome Res*. 2012;11:4517–25.
- Siuti N, Kelleher NL. Decoding protein modifications using top-down mass spectrometry. *Nat Methods*. 2007;4:817–21.
- Smith CA, O'maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G. METLIN: a metabolite mass spectral database. *Ther Drug Monit*. 2005;27:747–51.
- Spange S, Wagner T, Heinzel T, Krämer OH. Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *Int J Biochem Cell Biol*. 2009;41:185–98.

- Starheim KK, Gromyko D, Evjenth R, Rynningen A, Varhaug JE, Lillehaug JR, Arnesen T. Knockdown of human N α -terminal acetyltransferase complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization. *Mol Cell Biol.* 2009;29:3569–81.
- Sykes SM, Mellert HS, Holbert MA, Li K, Marmorstein R, Lane WS, McMahon SB. Acetylation of the p53 DNA-binding domain regulates apoptosis induction. *Mol Cell.* 2006;24:841–51.
- Tang H, Mechref Y, Novotny MV. Automated interpretation of MS/MS spectra of oligosaccharides. *Bioinformatics.* 2005;21 Suppl 1:i431–9.
- Tanner S, Shu HJ, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal Chem.* 2005;77:4626–39.
- Taus T, Kocher T, Pichler P, Paschke C, Schmidt A, Henrich C, Mechtler K. Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res.* 2011;10:5354–62.
- Tipton JD, Tran JC, Catherman AD, Ahlf DR, Durbin KR, Kelleher NL. Analysis of intact protein isoforms by mass spectrometry. *J Biol Chem.* 2011;286:25451–8.
- Tom CT, Martin BR. Fat chance! Getting a grip on a slippery modification. *ACS Chem Biol.* 2013;8:46–57.
- Tsur D, Tanner S, Zandi E, Bafna V, Pevzner PA. Identification of post-translational modifications by blind search of mass spectra. *Nat Biotechnol.* 2005;23:1562–7.
- Udeshi ND, Mani DR, Eisenhaure T, Mertins P, Jaffe JD, Clauser KR, Hacohen N, Carr SA. Methods for quantification of in vivo changes in protein ubiquitination following proteasome and deubiquitinase inhibition. *Mol Cell Proteomics.* 2012;11:148–59.
- Ujihara T, Sakurai I, Mizusawa N, Wada H. A method for analyzing lipid-modified proteins with mass spectrometry. *Anal Biochem.* 2008;374:429–31.
- Vakhrushev SY, Dadimov D, Peter-Katalinic J. Software platform for high-throughput glycomics. *Anal Chem.* 2009;81:3252–60.
- van Damme P, Arnesen T, Gevaert K. Protein alpha-N-acetylation studied by N-terminomics. *FEBS J.* 2011;278:3822–34.
- van Wijk SJ, Timmers HT. The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins. *FASEB J.* 2010;24:981–93.
- Varki A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology.* 1993;3:97–130.
- Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME. *Essentials of glycobiology.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2009.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, Mckusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, DI Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji R-R, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J-L, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, salzberg S, Shao W, Shue B, Sun J, Wang ZY, Wang A, Wang X, Wang J, Wei M-H, Wides R, Xiao C, Yan C, et al. The sequence of the human genome. *Science.* 2001;291:1304–51.
- von der Lieth CW, Freire AA, Blank D, Campbell MP, Ceroni A, Damerell DR, Dell A, Dwek RA, Ernst B, Fogh R, Frank M, Geyer H, Geyer R, Harrison MJ, Henrick K, Herget S, Hull WE, Ionides J, Joshi HJ, Kamerling JP, LeeFlang BR, Lutteke T, Lundborg M, Maass K, Merry A, Ranzinger R, rosen J, Royle L, Rudd PM, Schloissnig S, Stenutz R, Vranken WF, Widmalm G,

- Haslam SM. EUROCarbDB: an open-access platform for glycoinformatics. *Glycobiology*. 2011;21:493–502.
- Voutsadakis IA. Pathogenesis of colorectal carcinoma and therapeutic implications: the roles of the ubiquitin-proteasome system and Cox-2. *J Cell Mol Med*. 2007;11:252–85.
- Voutsadakis IA. Ubiquitination and the ubiquitin-proteasome system as regulators of transcription and transcription factors in epithelial mesenchymal transition of cancer. *Tumour Biol*. 2012;33:897–910.
- Wang W, Yang X, Kawai T, De Silanes IL, Mazan-Mamczarz K, Chen P, Chook YM, Quensel C, Köhler M, Gorospe M. AMP-activated protein kinase-regulated phosphorylation and acetylation of importin $\alpha 1$: involvement in the nuclear import of RNA-binding protein HuR. *J Biol Chem*. 2004;279:48376–88.
- Wang Z, Pandey A, Hart GW. Dynamic interplay between O-linked N-acetylglucosaminylation and glycogen synthase kinase-3-dependent phosphorylation. *Mol Cell Proteomics*. 2007;6:1365–79.
- Washburn MP, Wolters D, Yates 3rd JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*. 2001;19:242–7.
- Wimley WC, Creamer TP, White SH. Solvation energies of amino acid side chains and backbone in a family of host–guest pentapeptides†. *Biochemistry*. 1996;35:5109–24.
- Witze ES, Old WM, Resing KA, Ahn NG. Mapping protein post-translational modifications with mass spectrometry. *Nat Methods*. 2007;4:798–806.
- Woodworth A, Fiete D, Baenziger JU. Spatial and temporal regulation of tenascin-R glycosylation in the cerebellum. *J Biol Chem*. 2002;277:50941–7.
- Wotske M, WU Y, Wolters DA. Liquid chromatographic analysis and mass spectrometric identification of farnesylated peptides. *Anal Chem*. 2012;84:6848–55.
- Wu S-L, Huhmer AFR, Hao Z, Karger BL. On-line LC–MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. *J Proteome Res*. 2007;6:4230–44.
- Wuhrer M, Koeleman CAM, Hokke CH, Deelder AM. Protein glycosylation analyzed by normal-phase nano-liquid chromatography–mass spectrometry of glycopeptides. *Anal Chem*. 2004;77:886–94.
- Yang XJ, Seto E. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*. 2007;26:5310–18.
- Yang WH, Kim JE, Nam HW, Ju JW, Kim HS, Kim YS, Cho JW. Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. *Nat Cell Biol*. 2006;8:1074–83.
- Yang W, Di Vizio D, Kirchner M, Steen H, Freeman MR. Proteome scale characterization of human S-acylated proteins in lipid raft-enriched and non-raft membranes. *Mol Cell Proteomics*. 2010;9:54–70.
- Yuan H, Marmorstein R. Histone acetyltransferases: rising ancient counterparts to protein kinases. *Biopolymers*. 2013;99:98–111.
- Zeidan Q, Hart GW. The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways. *J Cell Sci*. 2010;123:13–22.
- Zeidman R, Jackson CS, Magee AI. Protein acyl thioesterases (Review). *Mol Membr Biol*. 2009;26:32–41.
- Zhang H, Aebersold R. Isolation of glycoproteins and identification of their N-linked glycosylation sites. In: *New and emerging proteomic techniques*. Totowa: Humana Press; 2006.
- Zhang H, Li X, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol*. 2003;21:660–6.
- Zhang H, Singh S, Reinhold VN. Congruent strategies for carbohydrate sequencing. 2. FragLib: an MSn spectral library. *Anal Chem*. 2005;77:6263–70.

- Zhang J, Planey SL, Ceballos C, Stevens Jr SM, Keay SK, Zacharias DA. Identification of CKAP4/p63 as a major substrate of the palmitoyl acyltransferase DHHC2, a putative tumor suppressor, using a novel proteomics method. *Mol Cell Proteomics*. 2008;7:1378–88.
- Zhang H, Guo T, Li X, Datta A, Park JE, Yang J, Lim SK, Tam JP, Sze SK. Simultaneous characterization of glyco- and phosphoproteomes of mouse brain membrane proteome with electrostatic repulsion hydrophilic interaction chromatography. *Mol Cell Proteomics*. 2010;9:635–47.
- Zozulya S, Stryer L. Calcium-myristoyl protein switch. *Proc Natl Acad Sci U S A*. 1992;89:11569–73.
- Zubarev RA, Kelleher NL, McLafferty FW. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc*. 1998;120:3265–6.
- Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem*. 2000;72:563–73.