Full text provided by www.sciencedirect.com

SCIENCE DIRECT®

Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes

Hui Zhang, Wei Yan and Ruedi Aebersold*

Quantitative proteome profiling using mass spectrometry and stable isotope dilution is being widely applied for the functional analysis of biological systems and for the detection of clinical, diagnostic or prognostic marker proteins. Because of the enormous complexity of proteomes, their comprehensive analysis is unlikely to be routinely achieved in the near future. However, in recent years, significant progress has been achieved focusing quantitative proteomic analyses on specific protein classes or subproteomes that are rich in biologically or clinically important information. Such projects typically combine the use of chemical probes that are specific for a targeted group of proteins and may contain stable isotope signatures for accurate quantification with automated tandem mass spectrometry and bioinformatics tools for data analysis. In this review, we summarize technical and conceptual advances in quantitative subproteome profiling based on tandem mass spectrometry and chemical probes.

Addresses
Institute for Systems Biology, 1441 N 34th Street, Seattle, Washington 98103-8904, USA
*e-mail: raebersold@systemsbiology.org

Current Opinion in Chemical Biology 2004, **8**:66–75

This review comes from a themed issue on
Proteomics and genomics
Edited by Michael Snyder and John Yates III

1367-5931/$ – see front matter
© 2003 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.cbpa.2003.12.001

Abbreviations

DTT	dithiothreitol
ICAT	isotope-coded affinity tag
IMAC	immobilized metal affinity chromatography
LC-MS/MS	liquid chromatography tandem mass spectrometry
MS	mass spectrometry
MS/MS	tandem MS
PNGaseF	peptide-N-glycosidase F
RP-HPLC	reverse-phase high-performance liquid chromatography

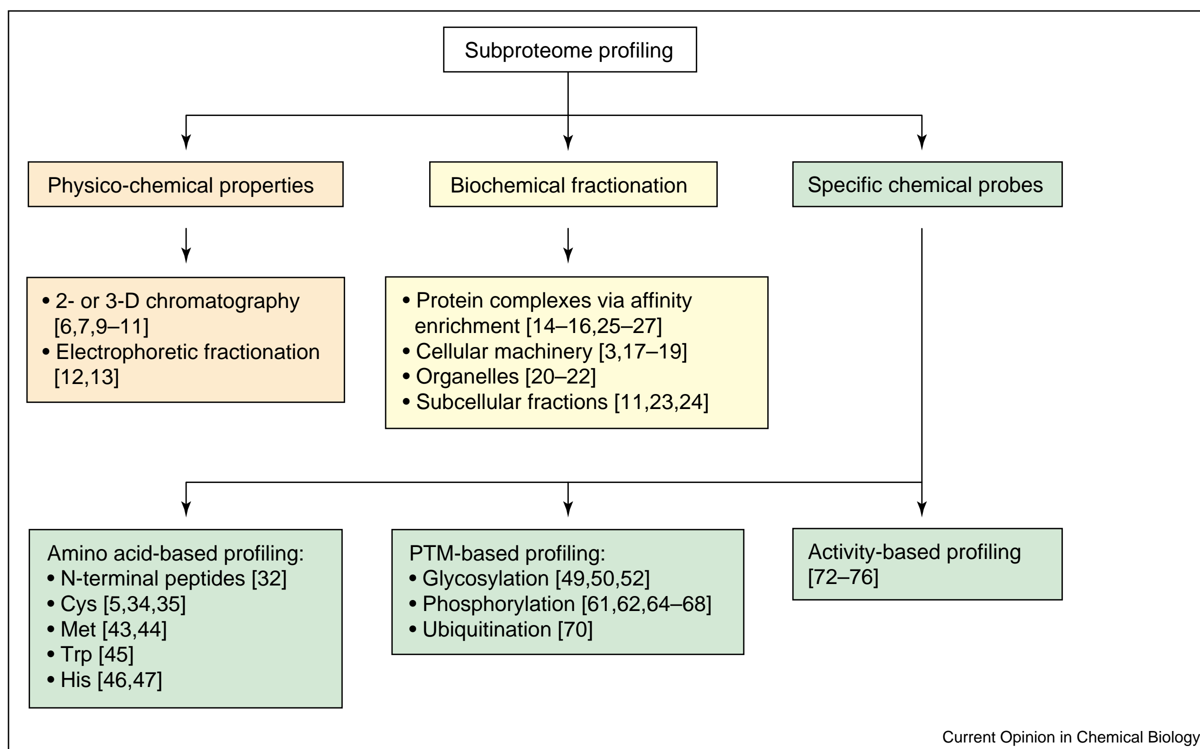
Introduction

Quantitative proteome profiling, defined as the systematic identification of the proteins in complex samples and the determination of their quantity or quantitative change, is an important component of the emerging science of systems biology. Quantitative protein profiling is expected to provide new functional insights into biological processes, facilitate the identification of diagnostic or prognostic disease markers, and contribute to the discovery of proteins as therapeutic targets.

Two methods have been most commonly used to generate quantitative profiles of complex protein mixtures. The first is a combination of two-dimensional gel electrophoresis and mass spectrometry (MS) [1,2]. The second is a more recently developed technique based on stable isotope tagging of proteins and automated liquid chromatography tandem MS (LC-MS/MS) analysis of peptides derived from complex protein mixtures [2–5]. Because of the 'top down' mode of protein detection and identification, neither method has successfully characterized the complete proteome of any species to date, although it appears that the LC-MS/MS-based methods are capable of identifying larger numbers of proteins efficiently [6–8]. Accepting the fact that complete proteome analysis will remain difficult using currently available methods, a 'divide and conquer' strategy has emerged to comprehensively analyze specific subsets of the proteome that are selectively isolated.

For the analysis of subproteomes by MS, three experimental approaches have been explored (summarized in Figure 1). In the first, proteins or peptides are fractionated according to their physico-chemical properties such as size, charge or hydropathy before their MS analysis. Specific implementations include two- or three-dimensional peptide chromatography [6,7,9•,10•,11] or electrophoretic fractionation of protein mixtures before their digestion and analysis by LC-MS/MS [12,13]. The second approach uses biochemical methods for the fractionation of functional protein complexes. Specific implementations include the isolation of protein complexes via affinity enrichment for one component in a protein complex [14,15•,16•], the isolation of large cellular machineries such as ribosomes [3], spliceosomes [17•,18•], and the nuclear pore complex [19], or organelles and subcellular fractions [11,20•,21,22•,23••,24] by density gradient centrifugation or equivalent established protocols. In each case, isolated samples can be analyzed by MS. In this approach, accurate quantification provided by stable isotope tagging has proven particularly advantageous for the discrimination of proteins specifically associated with the target population from non-specifically co-purified contaminants, as well as for the observation of quantitative changes in the population of target proteins [25••,26,27••]. The third approach uses specific chemical probes that can selectively tag and facilitate subsequent isolation of a target protein population such as phosphorylated or glycosylated proteins. Such isolated subproteomes are

Figure 1



Strategies for subproteome profiling, with key references.

also typically analyzed by LC-MS/MS. If the probes also provide a stable isotope tag, such analyses can also be performed quantitatively. These strategies have in common that they focus on the in-depth (ideally complete) analysis of subproteomes of rich biological context, thus minimizing repeated analysis of the abundant, but usually less biologically informative proteins. The following sections of this review focus on recent advances in the development and application of specific chemical probes for quantitative proteomics. For discussion of other strategies, the reader is referred to other recent reviews [28–30].

Probes and reactions selectively targeting specific amino acids

On average, tryptic digestion generates several dozen peptides per protein; profiling of complex proteomes by LC-MS/MS is therefore complicated by the very large number of redundant peptides. Theoretically, one unique peptide would be sufficient to unambiguously identify each parent protein. If such unique peptides could be isolated, the complexity of the samples for proteome profiling would be reduced by one to two orders of magnitude. Therefore, several strategies have been developed to target and isolate peptides containing unique (N- or C-terminal) or rare (Cys, Met, Trp, His etc.) amino acids. Table 1 illustrates the effect of such strategies if applied to human proteome. The values in the table are derived by calculating the total number of tryptic peptides from version 2.23 of the Human IPI database (International Protein Index human sequence database downloaded from the European Bioinformatics Institute; ftp://ftp.ebi.ac.uk/pub/databases/IPI/current/ipi.HUMAN.fasta.gz), which contains 52 816 protein entries, and the number of tryptic peptides that contain the target amino acid. In the following, we summarize recently published methods for the selective isolation of peptides from peptide mixtures by targeting unique or rare amino acids.

N-terminal peptides

Because each protein contains precisely one N-terminal and one C-terminal amino acid, it seems obvious to target these residues for efficient reduction of sample complexity. In practice, several factors complicate such experiments. First, the N-terminal amino and the C-terminal carboxyl groups are difficult to chemically distinguish from their counterparts in amino acid side chains. Second, the C-termini of some and the N-termini of a significant fraction of proteins are modified and therefore refractory to amino- and carboxyl-specific chemistries, respectively. Third, many proteins will yield terminal tryptic peptides that are not amenable for mass spectrometric analysis for reasons of inappropriate size or poor solubility [31••,32••].

Table 1

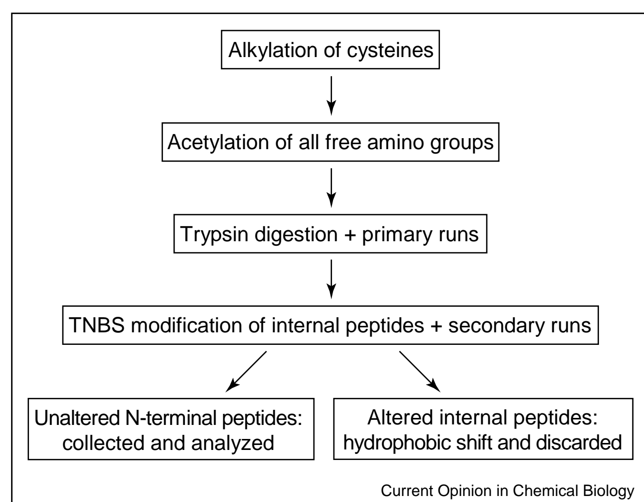
The calculated number of tryptic peptides containing the target amino acids and their representation coverage of the human proteome

	Total peptide number[a]	% of peptide	% of coverage
Total tryptic peptides[b]	892 584	100.0	100.0
N- or C-terminal peptides	52 816	5.9	100.0
Cys-containing peptides	237 111	26.6	96.1
Met-containing peptides	227 773	25.5	98.9
Trp-containing peptides	157 538	17.6	91.5
His-containing peptides	274 576	30.8	97.3
His- and Cys-containing peptides	95 238	10.7	73.2

[a] Human IPI database 2.23 with 52 816 total protein entries.
[b] The tryptic peptides are ended with Arg or Lys but not followed by proline.

Despite these potential problems, Gevaert and co-workers successfully developed a procedure for the enrichment of N-terminal peptides based on the concept of diagonal chromatography [32••,33]. In this method, analytes are separated twice in sequence by an identical separation method, in this case using reverse-phase-HPLC (RP-HPLC). Between the two runs, fractions collected from the first dimension are treated in a way that alters the retention time of the peptides containing the target amino acid. Each fraction is then analyzed in the second dimension using the same chromatographic conditions as in the first. Consequently, all modified peptides exhibit altered retention times and are thus known to contain the targeted amino acid whereas the remaining peptides elute from the column with the same retention time as in the first dimension. The significance of this method lies in the fact that it can be adopted to identify any subset of peptides that can be selectively modified. To specifically isolate N-terminal peptides, Gevaert and co-workers first acetylated all the amino groups before protein digestion. After separating the tryptic peptides in the first RP-HPLC dimension, the unblocked α-amino groups (i.e. the N-termini of the internal tryptic peptides) were modified using 2,4,6-trinitrobenzenesulfonic acid (TNBS) to induce a chromatographic shift towards the hydrophobic side. Therefore, labeled peptides were segregated from the unlabeled, predominantly N-terminal peptides during the second chromatographic dimension. The N-terminal peptides were then collected for subsequent identification via LC-MS/MS (Figure 2) [32••]. This approach significantly simplified the complex peptide mixtures, and also provided the exact N-terminal sequence of proteins. However, it was observed that some unmodified internal peptides, including those containing N-terminal proline or pyroglutamate residues, and peptides from abundant proteins, may not show a chromatographic shift and thus contaminate the N-terminal peptide population [32••].

Figure 2

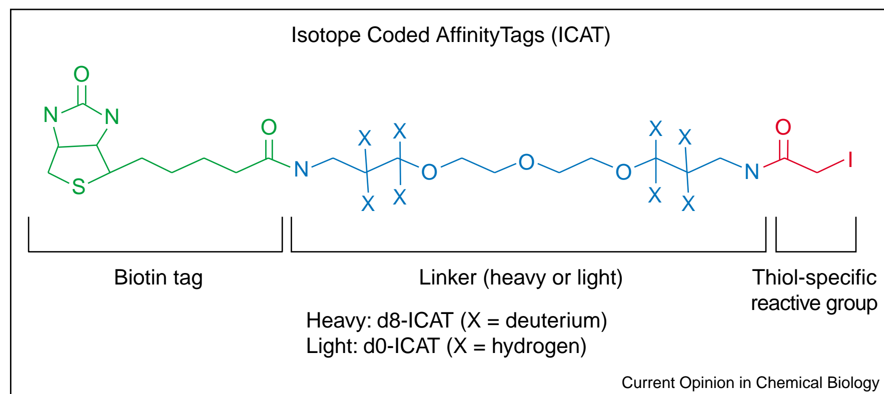


Schematic diagram of steps for N-terminal peptide sorting. TNBS, 2,4,6-trinitrobenzenesulfonic acid.

Cysteine-containing peptides

Cysteine residues represent a good target to reduce the complexity of digests from large proteomes or other complex protein samples, because the 26.6% of the total human tryptic peptides that contain cysteine represent 96.1% of all human proteins (Table 1). Furthermore, there are several chemistries available to target the SH group on the side chain of reduced cysteine. Therefore cysteine became one of the earliest targets for amino-acid-based peptide selection [5,34] and the isotope-coded affinity tag (ICAT) reagent method [5] has become widely used for quantitative proteome profiling. The ICAT reagents include a cysteine-reactive group, an isotopically light or heavy linker group and an affinity tag (biotin) (Figure 3). The light and heavy reagents are used to label the cysteine residues from proteins of two different sources. The labeled proteins are then combined and proteolyzed with trypsin. Cysteine-containing peptides are isolated using avidin affinity chromatography and subsequently identified, and quantified by MS. Because only the cysteine-containing peptides are analyzed in the mass spectrometer, more proteins can be identified within a specific analysis time than when all tryptic peptides are analyzed. Because mass spectrometers typically select the precursor ions with the highest signal intensity first, a larger number

Figure 3



The structure of ICAT reagents, which comprise a cysteine-reactive group (red), a linker containing either heavy or light isotopes (blue) and a binding affinity tag (green).

of proteins identified automatically translates into the identification of lower-abundance proteins [10•].

More recently, a solid-phase variant of the ICAT method has been developed in which a small stable isotope tag is attached to cysteine after capture and release of cysteine-containing peptides [35••]. Additionally, second-generation ICAT reagents have become commercially available that contain an acid-cleavable group between the linker and the biotin group. As a result, peptides labeled with this reagent contain a smaller attachment at the cysteine side chain, which reduces collision energy lost in fragmentation of the ICAT linker [12,36]. Several studies have successfully used the ICAT reagent method to answer a broad range of biological questions. These include the analysis of Myc oncoprotein cellular functions [26,37], microsomal proteins in response to pharmacologically induced differentiation in human myeloid leukemia cells (HL-60) [11], the analysis of multiprotein complexes [25••], the identification of T cell receptor/CD28 cross-linking induced changes in the protein composition of lipid rafts isolated from Jurkat T cells [38], the discovery of new cellular pathways [39,40••] and their interconnectivity in cells [41]. Furthermore, inflammatory proteins in tendinitis [42] and candidate drug targets [36] have been identified.

Methionine-containing peptides

Two approaches have been reported for the analysis of methionine-containing peptides. The first uses commercially available methionine-specific beads to covalently attach methionine-containing peptides via bromoacetyl functional groups to a solid support [43]. After removal of the unbound peptides, the methionine-containing peptides are released from the support with β-mercaptoethanol, and analyzed by MS with similar benefits for complex mixture analysis as observed with the cysteine-capture method.

The second approach is again based on the concept of diagonal chromatography as described above [32••,33]. In this case, methionine side chains are oxidized between chromatographic runs, resulting in increased polarity and therefore earlier RP-HPLC elution of the oxidized peptides in the second dimension away from the bulk of unmodified peptides [44••]. Both technologies targeting the methionine-containing peptides are limited to protein identification only at this stage as neither method introduces a stable isotope tag for accurate quantification. However, both methods could be implemented as quantitative techniques if they were combined with isotope labeling techniques. A drawback of such methionine enrichment approaches is that some methionine residues typically become oxidized in cells or during sample preparation, which may lead to significantly reduced and potentially variable yields.

Tryptophan-containing peptides

A method for targeting tryptophan-containing peptides has been recently developed where tryptophan residues are labeled with 2-nitrobenzenesulfenyl chloride (NBSCl) before analysis by MS/MS [45•]. By using two isotopic forms of the reagent, labeled peptide pairs with a mass difference of 6 Da can be generated that enable quantitative comparison of proteins isolated from related samples. Tryptophan-containing peptides were enriched by a Sephadex column based on the hydrophobic property of peptides containing the modified tryptophan. Although somewhat less selective, this method complements the cysteine- and methionine-specific chemistries described above.

Histidine-containing peptides

Histidine-containing peptides from complex peptide mixtures can be enriched by immobilized metal affinity chromatography (IMAC) loaded with Cu^{2+} [46]. To

minimize non-specific interactions of the IMAC resin with peptides containing cysteine, tryptophan and carboxyl groups, samples were reduced, alkylated and acetylated before affinity capture of the histidine-containing peptides by IMAC columns. To further simplify the peptide complexity, peptides containing both cysteine and histidine have also been enriched and analyzed [47]. Here, cysteine-containing peptides were selected first by covalent chromatography using thiol disulfide exchange. Following the release of cysteine-containing peptides from the covalent chromatography with reductive cleavage, histidine-containing peptides were subsequently captured by passage through an IMAC column loaded with Cu^{2+}. Selection of peptides containing both cysteine and histidine residues further simplified the peptide complexity at the cost of some loss of protein coverage (Table 1).

Subproteome profiling based on post-translational modifications

The residue-based selection and quantitative analysis of one or a few representative peptides from each protein can efficiently profile whole proteomes or subproteomes. However, the reduced sequence coverage caused by the selective isolation of a subset of peptides usually precludes the identification of important structural features such as post translational modifications, sequence variations induced by differential splicing and other forms of protein processing. To address this issue, and as an extension to the concept of selective isolation of subsets of peptides using chemical probes, the following methods have been developed that aim to selectively target post-translationally modified peptides.

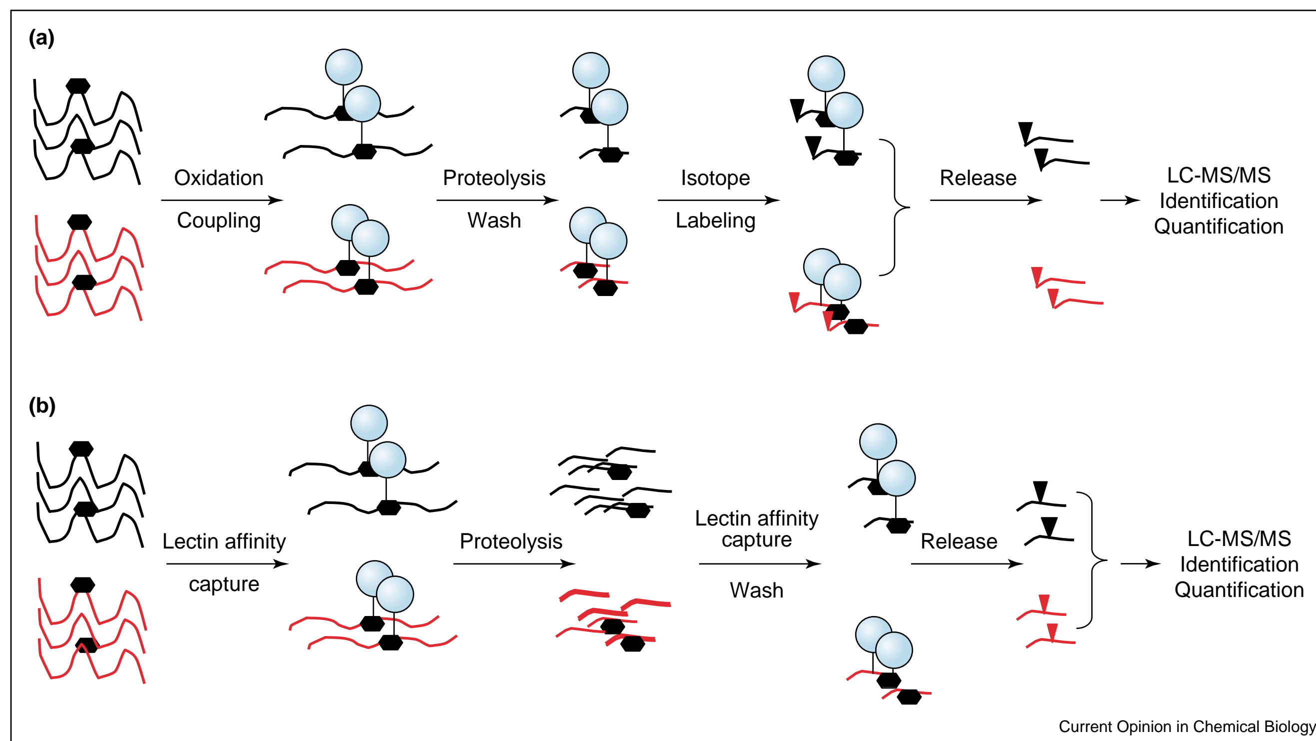
Glycosylation

Protein glycosylation has long been recognized as a common post-translational modification and has been increasingly recognized as one of the most prominent biochemical alterations associated with malignant transformation and tumorgenesis. Typically, carbohydrates are linked to the side chains of serine or threonine residues (O-linked glycosylation) or of asparagine residues (N-linked glycosylation). N-linked glycosylation sites generally fall into the N-X-S/T sequence motif in which X denotes any amino acid except proline [48]. N-linked glycosylation is prevalent in proteins on the extracellular side of the plasma membrane, secreted proteins, and proteins retained in body fluids (blood serum, cerebrospinal fluid, urine, breast milk, saliva, lung lavage fluid, pancreatic juice, etc.). Many clinical biomarkers and therapeutic targets are glycoproteins. However, after years of efforts by many researchers in the field, the current Protein Information Resources – Protein Sequence Database (http://pir.georgetown.edu/pirwww/search/textpsd.shtml) contains only 172 experimentally confirmed human glycoproteins [49••].

Recently, however, two new methods for the identification of N-linked glycopeptides in complex biological samples have been reported [49••,50••]. As shown in Figure 4, both methods first immobilize glycopeptides on a solid support, then the N-linked glycopeptides are released by peptide-N-glycosidase. MS is then used to identify the N-linked glycosylation sites and quantify the relative abundance of glycopeptides using isotope-coded tags. The difference of the two approaches lies in the mechanisms by which the glycopeptides are captured to solid support and labeled with isotope tags. In the first approach, covalent conjugation of glycoproteins to a solid support via hydrazide chemistry is used (Figure 4a) [50••]. Periodate oxidation converts the *cis*-diol groups of carbohydrates to aldehydes, which then form covalent hydrazone bonds with hydrazide groups immobilized on the solid support. After the non-glycosylated proteins are washed away, the immobilized glycoproteins are proteolyzed on the solid support. Non-glycosylated peptides are again washed away while the glycosylated peptides remain on the solid support. The α-amino groups of the immobilized glycopeptides are next labeled with isotopically light (d0, containing no deuterium) or heavy (d4, containing four deuterium atoms) forms of succinic anhydride after the ε-amino groups of lysine have been converted to homoarginine. The formerly N-linked glycosylated peptides are finally released from the solid phase using peptide-N-glycosidase F (PNGaseF). This PNGaseF treatment also results in the conversion of the glycosylated asparagines to aspartic acid, thus generating one unit mass shift at the formerly glycosylated asparigine residues, that is, in a high resolution mass spectrometer, diagnostic for the glycosylation site. Therefore, in a single analysis, the method identifies N-glycosylated proteins, the site(s) of N-glycosylation, and the relative quantity of the identified glycopeptides via stable isotope tagging. The method has also been successfully applied to the analysis of cell surface proteins and for serum proteome profiling. Interestingly, albumin, the most abundant serum protein, does not appear to contain any N-linked glycosylation sites and is therefore transparent to the method, allowing for more efficient identification and quantification of the lower abundance glycoproteins [50••].

The second approach is based on lectin-column-mediated affinity capture (Figure 4b) [49••]. Here, the glycoproteins are first purified from a biological protein mixture using lectin-affinity chromatography. The recovered glycoproteins are digested with trypsin and glycopeptides are captured again using the same lectin column. The N-linked glycopeptides are then cleaved with PNGaseF in ^{18}O water to incorporate the glycosylation-site-specific tag. Using this method, 250 N-linked glycoproteins with 400 unique N-glycosylation sites were identified in *Caenorhabditis elegans*, with an average of 1.6 glycosylation sites per protein. The two unit mass shift

Figure 4



Schematic diagram of quantitative analysis of N-linked glycopeptides using **(a)** hydrazide chemistry or **(b)** lectin affinity capture. See text for full details.

introduced by PNGaseF-mediated reaction in $H_2{}^{16}O$/$H_2{}^{18}O$ can be used for quantitative profiling of glycoproteins by isotope dilution. However, the isotope distributions of the ^{16}O- and ^{18}O-tagged peptides partially overlap because of natural isotope distributions, making it difficult to achieve accurate quantification.

One unique feature of the first approach is that both N-linked and O-linked glycoproteins are conjugated to solid support via covalent linkage [50••]. PNGaseF releases the N-glycosylated peptides from the respective solid supports specifically. To successfully release O-linked oligosaccharides, it is necessary to use a panel of exoglycosidases to sequentially remove monosaccharides until only the Galβ1,3GalNAc core remains attached to the serine or threonine residue. The core can then be released by O-glycosidase. Because not all O-linked oligosaccharides contain this core structure, a chemical method, such as β-elimination, may be more general and effective for the release of the formerly O-linked glycosylated peptides.

O-GlcNAc is a dynamic nucleocytoplasmic post translational modification that is more analogous to phosphorylation than to classical complex O-glycosylation [51]. Mild β-elimination followed by replacement of the carbohydrate with dithiothreitol (DTT) through the Michael addition has been shown to be successful for the identification of protein O-GlcNAc modification sites [52••]. The introduced DTT group can be used for affinity-enrichment of the tagged peptides using an activated thiol-Sepharose column before their analysis by LC-MS/MS. In addition, if two different isotopic compositions of DTT are used, O-GlcNAc-linked peptides from related samples can also be quantified. Because β-elimination followed by Michael addition has also been used to determine the sites of serine and threonine phosphorylation, it is important to distinguish in such experiments between O-phosphate and O-GlcNAc modifications. This can be achieved using modification-specific antibodies such as anti-O-GlcNAc antibodies or phospho-specific antibodies and/or by the use of enzymes that remove any phosphate or O-GlcNAc before the chemical reactions are carried out [52••].

Phosphorylation

Reversible protein phosphorylation is a common post-translational modification that affects a broad range of biological functions and processes. Most commonly, phosphorylation occurs at the side chains of serine, threonine or tyrosine residues, although other types of amino acids including histidine and glutamic acid can also be

phosphorylated. For the analysis of protein phosphorylation, three general strategies have emerged. The first is based on the use of specific MS techniques for the detection and identification of phosphorylated peptides in simple mixtures [31••,53]. These methods do not rely on the use of selective probes and thus are beyond the scope of this review. The second strategy is based on the selective isolation of phosphorylated proteins and the third on the selective isolation of phosphorylated peptides.

Affinity purifications of phosphoproteins using antibodies for a particular type of phosphorylated amino acid or for specific amino acid sequences containing phosphorylated amino acids have been used in a wide range of biological studies. The availability of antibodies that specifically recognize phosphotyrosine has enabled proteomic profiling of tyrosine-phosphorylated proteins based on immuno-affinity purification followed by sequence identification by MS [54,55]. However, greater than 90% of protein phosphorylation occurs at serine or threonine residues and immuno-affinity purification of phosphoserine or phosphothreonine proteins has been less successful due to the lack of high affinity and broadly reactive phosphoserine- or phosphothreonine-specific antibodies. Recently, two groups have reported the successful identification of novel protein kinase A [56•] or Akt substrate proteins [57•] using recently developed antibodies broadly reactive to a consensus recognition motif of a respective kinase, followed by MS/MS [58•]. However, as the enrichment is at the protein level in these studies, many non-phosphorylated proteins are co-purified, thus complicating the determination of truly phosphorylated proteins. Furthermore, even if a phosphoprotein is correctly identified, it is often difficult to identify the phosphorylated residue.

To gain more direct access to the identification of phosphorylation sites, several methods for the selective enrichment of phosphopeptides before their analysis by LC-MS/MS have been used. First, phosphopeptides are enriched by IMAC. IMAC uses metal ions such as Fe^{3+} or Ga^{3+} that have affinity for phosphorylated residues [59,60]. Because the same metal ions also have affinity for other groups, including carboxyls, however, this application of IMAC has been limited. Recently, the selectivity of IMAC for phosphopeptides has been improved by capping peptide carboxyl groups via methyl-esterification. This methodological modification has led to the identification of hundreds of phosphopeptides in a yeast cell lysate [61•]. However, another recent study found that methyl-esterification did not decrease the binding of unphosphorylated peptides to the IMAC resin; rather, it increased the specificity of elution of phosphopeptides with phosphate buffer [62]. In the second strategy, chemical modifications combined with affinity purification were used to isolate phosphopeptides [63]. Chemical derivatization of phosphoamino acids has been achieved by phosphoamidate chemistry [64], or by a β-elimination and Michael addition reaction of the phosphorylation site, thus attaching an affinity tag at the formerly phosphorylated residue [65–67]. The advantage of the phosphoamidate chemistry is that it allows the mapping of phosphorylation sites at serine, threonine and tyrosine. However, the procedure is relatively complicated and the sample yields are low. The approach based on β-elimination and Michael addition reaction is limited to phosphorylated serine or threonine residues. In addition, it is not specific to phosphorylation as other modifications to serines and threonines, such as O-GlcNAc modification and O-linked oligosaccharide modification are also susceptible to β-elimination and thus targeted by this method. Other approaches, such as modification-specific antibodies, enzymatic reaction, in combination with quantitative proteomic technologies can be used to distinguish the different modifications occurring in the serine and threonine sites. Recently, the same β-elimination and addition reactions were also used to create lysine-specific protease sites at the sites of serine or threonine phosphorylation [68••]. Dehydroalanine or β-methyldehydroalanine generated after β-elimination from phosphoserine or phosphothreonine residues, respectively, were transformed to lysine analogs (aminoethylcysteine and β-methylaminoethylcysteine) after reaction with cysteamine. Proteases that recognize lysine (e.g. trypsin, Lys-C and lysyl endopeptidase) then cleave proteins at these residues, thus mapping the sites of phosphorylation. The advantage of this site-specific modification followed by proteolytic digestion is that it allows identification of serine and threonine phosphorylation sites directly from the cleavage pattern without the necessity of peptide sequencing. In addition, it makes interpretation of the MS/MS spectra easier because the C-terminal residue is always the phosphorylated residue. Selective capture and modification of phosphopeptides using cysteamine on a solid support could in the future facilitate phosphopeptide enrichment and identification in a single step.

Ubiquitination

Protein ubiquitination plays an important role in regulation of protein degradation. During the ubiquitination process, ubiquitin is covalently conjugated to the ε-amino group of lysine residues of target proteins. Tryptic digestion of ubiquitinated proteins produces a signature peptide at the ubiquitination site containing a di-glycine remnant (G-G) that can be detected by mass spectrometers [69]. Gygi *et al.* have used His-tagged ubiquitin to capture ubiquitinated proteins from yeast cells and searched for the G-G-containing peptide during MS/MS analysis. They successfully identified 1075 proteins and 110 ubiquitination sites that were present in 72 ubiquitin–protein conjugates [70••]. This approach, together with the quantitative proteomic technologies, will open up the study of protein modification

by ubiquitination and other ubiquitin-like protein modifiers in a cell state- and protein complex-dependent manner [71].

Probes and reactions for activity-based protein profiling

Profiling the activities of certain classes of proteins using activity-based chemical probes is a rapidly evolving area of proteomics. Activity-based profiling uses chemical probes to label specific target enzyme superfamilies on the basis of functional activity. The activity-based probe is typically composed of reactive group that binds and covalently modifies the active sites of all the members of a given enzyme class, and a chemical tag for the detection or isolation of bound enzymes. These probes have been developed from the extensive understanding of the catalytic mechanisms of the enzyme family studied. The strategy has been successfully implemented for serine hydrolases [72], cysteine proteases [73] and protein phosphatases [74•]. Activity-based chemical probes were also used in competitive profiling of a library of candidate serine hydrolase inhibitors in complex proteomes to select for reversible enzyme inhibitors [75••]. However, for those enzymes that lack cognate affinity labels or for which the catalytic mechanism is not well understood, it is not possible to rationally design activity-based proteomics probes. In a recent study, Cravatt's group used a library with a sulfonate ester chemotype to screen complex proteomes for activity-dependent protein activities [76••]. The screen resulted in the labeling of at least six mechanistically distinct enzyme classes, and none of the sulfonate-labeled enzymes represented targets that had been previously identified using specific proteomics probes.

Conclusions

Recently, mass spectrometric technologies to detect and sequence large numbers of peptides and informatics tools to analyze the huge amounts of data created by such experiments have progressed dramatically. Apart from analyzing entire proteomes, the idea of thoroughly analyzing subproteomes of particular biological interest has been steadily gaining attention because with such a strategy, the sample complexity and therefore the analytical challenge are reduced, while the biological information obtained has been enriched in relevance for the system studied. The combination of chemical probes for the selective isolation of subproteomes and MS/MS for their analysis has proven particularly successful in this context. Furthermore, the use of such chemical probes has also facilitated the site-specific introduction of stable isotope tags that provide a basis for accurate quantification of the targeted analytes by MS. As the approach is only limited by the creativity of the chemists that apply it, we can expect that numerous additional probes will be developed that will provide yet more insights into the structure, function and control of biological processes.

Acknowledgements

We thank Jimmy Eng for Table 1, and Julian Watts, Sharon Chen and Daniel Martin for proof reading. This work was supported in part by grants from the National Cancer Institute (R33, CA93302), by federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Proteomics Initiative (N01-HV-28179), and National Institute for Drug Abuse (1P30DA01562501) of the National Institutes of Health and by a sponsored research agreement from MacroGenics. The ISB is supported by a generous gift from Merck and CO.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Patterson SD: **Proteomics: the industrialization of protein chemistry**. *Curr Opin Biotechnol* 2000, **11**:413-418.

2. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT: **Accurate quantitation of protein expression and site-specific phosphorylation**. *Proc Natl Acad Sci USA* 1999, **96**:6591-6596.

3. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR III: **Direct analysis of protein complexes using mass spectrometry**. *Nat Biotechnol* 1999, **17**:676-682.

4. Veenstra TD, Martinovic S, Anderson GA, Pasa-Tolic L, Smith RD: **Proteome analysis using selective incorporation of isotopically labeled amino acids**. *J Am Soc Mass Spectrom* 2000, **11**:78-82.

5. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags**. *Nat Biotechnol* 1999, **17**:994-999.

6. Wolters DA, Washburn MP, Yates JR III: **An automated multidimensional protein identification technology for shotgun proteomics**. *Anal Chem* 2001, **73**:5683-5690.

7. Washburn MP, Wolters D, Yates JR III: **Large-scale analysis of the yeast proteome by multidimensional protein identification technology**. *Nat Biotechnol* 2001, **19**:242-247.

8. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R: **Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology**. *Proc Natl Acad Sci USA* 2000, **97**:9390-9395.

9. • Adkins JN, Varnum SM, Auberry KJ, Moore RJ, Angell NH, Smith RD, Springer DL, Pounds JG: **Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry**. *Mol Cell Proteomics* 2002, **1**:947-955.

This is a good example of multidimensional peptide chromatography based on physico-chemical properties to profile a human subproteome from blood serum.

10. • Gygi SP, Rist B, Griffin TJ, Eng J, Aebersold R: **Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags**. *J Proteome Res* 2002, **1**:47-54.

This paper describes classification of the proteome using multidimensional chromatography using ICAT-based quantification.

11. Han DK, Eng J, Zhou H, Aebersold R: **Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry**. *Nat Biotechnol* 2001, **19**:946-951.

12. Li J, Steen H, Gygi SP: **Protein profiling with cleavable isotope coded affinity tag (cICAT) reagents: the yeast salinity stress response**. *Mol Cell Proteomics* 2003, **2**:1198-1204.

13. Righetti PG, Castagna A, Herbert B, Reymond F, Rossier JS: **Prefractionation techniques in proteome analysis**. *Proteomics* 2003, **3**:1397-1407.

14. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B: **A generic protein purification method for protein complex characterization and proteome exploration**. *Nat Biotechnol* 1999, **17**:1030-1032.

15. • Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM *et al.*: **Functional**

organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002, **415**:141-147.
This paper describes identification of a functional protein complex using tandem affinity tags and MS to build up a protein interaction network.

16. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A,
• Taylor P, Bennett K, Boutilier K *et al.*: **Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry**. *Nature* 2002, **415**:180-183.
This paper reports another example of profiling protein complexes.

17. Rappsilber J, Ryder U, Lamond AI, Mann M: **Large-scale**
• **proteomic analysis of the human spliceosome**. *Genome Res* 2002, **12**:1231-1245.
This is a good example of proteomic analysis on cellular machinery.

18. Zhou Z, Licklider LJ, Gygi SP, Reed R: **Comprehensive proteomic**
• **analysis of the human spliceosome**. *Nature* 2002, **419**:182-185.
This paper provides an example of a subproteome study on the human spliceosome complex.

19. Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, Chait BT: **The yeast nuclear pore complex: composition, architecture, and transport mechanism**. *J Cell Biol* 2000, **148**:635-651.

20. Andersen JS, Lyon CE, Fox AH, Leung AK, Lam YW, Steen H,
• Mann M, Lamond AI: **Directed proteomic analysis of the human nucleolus**. *Curr Biol* 2002, **12**:1-11.
This paper gives an example of MS analysis on a subcellular fraction.

21. Fountoulakis M, Berndt P, Langen H, Suter L: **The rat liver mitochondrial proteins**. *Electrophoresis* 2002, **23**:311-328.

22. Yi EC, Marelli M, Lee H, Purvine SO, Aebersold R, Aitchison JD,
• Goodlett DR: **Approaching complete peroxisome characterization by gas-phase fractionation**. *Electrophoresis* 2002, **23**:3205-3216.
This paper describes improved coverage of peroxisome profiling using a gas phase fractionation technique.

23. Wu CC, MacCoss MJ, Howell KE, Yates JR: **A method for the**
•• **comprehensive proteomic analysis of membrane proteins**. *Nat Biotechnol* 2003, **21**:532-538.
This paper discusses technological development on proteomic analysis, particularly on membrane proteins.

24. Bergquist J, Gobom J, Blomberg A, Roepstorff P, Ekman R: **Identification of nuclei associated proteins by 2D-gel electrophoresis and mass spectrometry**. *J Neurosci Methods* 2001, **109**:3-11.

25. Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J,
•• Aebersold R: **The study of macromolecular complexes by quantitative proteomics**. *Nat Genet* 2003, **33**:349-355.
This is a good example of identification of functional protein complexes and discrimination of real components of a complex from contaminant proteins using quantitative proteomics.

26. Shiio Y, Eisenman RN, Yi EC, Donohoe S, Goodlett DR, Aebersold R: **Quantitative proteomic analysis of chromatin-associated factors**. *J Am Soc Mass Spectrom* 2003, **14**:696-703.

27. Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M:
•• **A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling**. *Nat Biotechnol* 2003, **21**:315-318.
This paper reports another MS-based approach to identify protein complexes from background binding using quantitative proteomics by differentially label proteins in EGF-stimulated versus unstimulated cells using stable isotopic amino acids in cell culture (SILAC).

28. Aebersold R, Mann M: **Mass spectrometry-based proteomics**. *Nature* 2003, **422**:198-207.

29. Tyers M, Mann M: **From genomics to proteomics**. *Nature* 2003, **422**:193-197.

30. Sechi S, Oda Y: **Quantitative proteomics using mass spectrometry**. *Curr Opin Chem Biol* 2003, **7**:70-77.

31. MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM,
•• Tasto JJ, Gould KL, Wolters D, Washburn M, Weiss A *et al.*: **Shotgun identification of protein modifications from protein complexes and lens tissue**. *Proc Natl Acad Sci USA* 2002, **99**:7900-7905.
This paper reports on proteomics analysis of a complex protein mixture and post-translational modification.

32. Gevaert K, Goethals M, Martens L, Van Damme J, Staes A,
•• Thomas GR, Vandekerckhove J: **Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides**. *Nat Biotechnol* 2003, **21**:566-569.
This paper describes a method to enrich N-terminal peptides using the diagonal chromatography approach.

33. Cruickshank WH, Malchy BL, Kaplan H: **Diagonal chromatography for the selective purification of tyrosyl peptides**. *Can J Biochem* 1974, **52**:1013-1017.

34. Spahr CS, Susin SA, Bures EJ, Robinson JH, Davis MT, McGinley MD, Kroemer G, Patterson SD: **Simplification of complex peptide mixtures for proteomic analysis: reversible biotinylation of cysteinyl peptides**. *Electrophoresis* 2000, **21**:1635-1650.

35. Zhou H, Ranish JA, Watts JD, Aebersold R: **Quantitative proteome**
•• **analysis by solid-phase isotope tagging and mass spectrometry**. *Nat Biotechnol* 2002, **20**:512-515.
A solid-phase-based stable isotope tagging method is discussed.

36. Oda Y, Owa T, Sato T, Boucher B, Daniels S, Yamanaka H, Shinohara Y, Yokoi A, Kuromitsu J, Nagasu T: **Quantitative chemical proteomics for identifying candidate drug targets**. *Anal Chem* 2003, **75**:2159-2165.

37. Shiio Y, Donohoe S, Yi EC, Goodlett DR, Aebersold R, Eisenman RN: **Quantitative proteomic analysis of Myc oncoprotein function**. *EMBO J* 2002, **21**:5088-5096.

38. Von Haller PD, Yi E, Donohoe S, Vaughn K, Keller A, Nesvizhskii AI, Eng J, Li XJ, Goodlett DR, Aebersold R *et al.*: **The application of new software tools to quantitative protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry: II. Evaluation of tandem mass spectrometry methodologies for large-scale protein analysis, and the application of statistical tools for data analysis and interpretation**. *Mol Cell Proteomics* 2003, **2**:428-442.

39. Baliga NS, Pan M, Goo YA, Yi EC, Goodlett DR, Dimitrov K, Shannon P, Aebersold R, Ng WV, Hood L: **Coordinate regulation of energy transduction modules in *Halobacterium* sp. analyzed by a global systems approach**. *Proc Natl Acad Sci USA* 2002, **99**:14913-14918.

40. Guina T, Purvine SO, Yi EC, Eng J, Goodlett DR, Aebersold R,
•• Miller SI: **Quantitative proteomic analysis indicates increased synthesis of a quinolone by *Pseudomonas aeruginosa* isolates from cystic fibrosis airways**. *Proc Natl Acad Sci USA* 2003, **100**:2771-2776.
This is a good example of the application of quantitative proteomics using ICAT reagents to address biological questions.

41. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L: **Integrated genomic and proteomic analyses of a systematically perturbed metabolic network**. *Science* 2001, **292**:929-934.

42. Harris RD, Nindl G, Balcavage WX, Weiner W, Johnson MT: **Use of proteomics methodology to evaluate inflammatory protein expression in tendinitis**. *Biomed Sci Instrum* 2003, **39**:493-499.

43. Weinberger SR, Viner RI, Ho P: **Tagless extraction-retentate chromatography: a new global protein digestion strategy for monitoring differential protein expression**. *Electrophoresis* 2002, **23**:3182-3192.

44. Gevaert K, Van Damme J, Goethals M, Thomas GR, Hoorelbeke B,
•• Demol H, Martens L, Puype M, Staes A, Vandekerckhove J: **Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: identification of more than 800 *Escherichia coli* proteins**. *Mol Cell Proteomics* 2002, **1**:896-903.
This paper reports on a useful method to study methionine-containing peptides based on the diagonal chromatography concept.

45. Kuyama H, Watanabe M, Toda C, Ando E, Tanaka K, Nishimura O:
• **An approach to quantitative proteome analysis by labeling tryptophan residues**. *Rapid Commun Mass Spectrom* 2003, **17**:1642-1650.
A method for targeting tryptophan-containing peptides has been proposed.

46. Ji J, Chakraborty A, Geng M, Zhang X, Amini A, Bina M, Regnier F: **Strategy for qualitative and quantitative analysis in proteomics**

based on signature peptides. *J Chromatogr B Biomed Sci Appl* 2000, **745**:197-210.

47. Wang S, Zhang X, Regnier FE: **Quantitative proteomics strategy involving the selection of peptides containing both cysteine and histidine from tryptic digests of cell lysates**. *J Chromatogr A* 2002, **949**:153-162.

48. Bause E: **Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes**. *Biochem J* 1983, **209**:331-336.

49. •• Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, Kasai K, Takahashi N, Isobe T: **Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins**. *Nat Biotechnol* 2003, **21**:667-672.

This paper describes MS analysis of protein glycosylations using the combined techniques of lectin affinity purification and isotope-coded tagging.

50. •• Zhang H, Li XJ, Martin DB, Aebersold R: **Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry**. *Nat Biotechnol* 2003, **21**:660-666.

This paper describes a method to identify and quantify N-linked glycoproteins. It also applies the method to enrich cell surface proteins and efficiently profile serum proteins by removing albumin from serum samples.

51. Wells L, Hart GW: **O-GlcNAc turns twenty: functional implications for post-translational modification of nuclear and cytosolic proteins with a sugar**. *FEBS Lett* 2003, **546**:154-158.

52. •• Wells L, Vosseller K, Cole RN, Cronshaw JM, Matunis MJ, Hart GW: **Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications**. *Mol Cell Proteomics* 2002, **1**:791-804.

Analysis of O-GlcNAc modification using β-elimination and Michael addition reaction is described. Also described is a strategy to distinguish O-GlcNAc modification at serine or threonine from phosphorylation modification using this approach.

53. Aebersold R, Goodlett DR: **Mass spectrometry in proteomics**. *Chem Rev* 2001, **101**:269-295.

54. Pandey A, Podtelejnikov AV, Blagoev B, Bustelo XR, Mann M, Lodish HF: **Analysis of receptor signaling pathways by mass spectrometry: identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors**. *Proc Natl Acad Sci USA* 2000, **97**:179-184.

55. Yeung YG, Wang Y, Einstein DB, Lee PS, Stanley ER: **Colony-stimulating factor-1 stimulates the formation of multimeric cytosolic complexes of signaling proteins and cytoskeletal components in macrophages**. *J Biol Chem* 1998, **273**:17128-17137.

56. • Gronborg M, Kristiansen TZ, Stensballe A, Andersen JS, Ohara O, Mann M, Jensen ON, Pandey A: **A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate**. *Mol Cell Proteomics* 2002, **1**:517-527.

This is an example of identification of phospho-Ser/Thr proteins based on phospho-specific antibodies.

57. • Kane S, Sano H, Liu SC, Asara JM, Lane WS, Garner CC, Lienhard GE: **A method to identify serine kinase substrates. Akt phosphorylates a novel adipocyte protein with a Rab GTPase-activating protein (GAP) domain**. *J Biol Chem* 2002, **277**:22115-22118.

This is another example of profiling serine phosphorylation using phospho-specific antibodies and MS.

58. • Zhang H, Zha X, Tan Y, Hornbeck PV, Mastrangelo AJ, Alessi DR, Polakiewicz RD, Comb MJ: **Phosphoprotein analysis using antibodies broadly reactive against phosphorylated motifs**. *J Biol Chem* 2002, **277**:39379-39387.

This paper describes the development of phospho-specific antibodies that are broadly reactive to the consensus phosphorylation motif of certain kinases.

59. Andersson L, Porath J: **Isolation of phosphoproteins by immobilized metal (Fe^{3+}) affinity chromatography**. *Anal Biochem* 1986, **154**:250-254.

60. Posewitz MC, Tempst P: **Immobilized gallium(III) affinity chromatography of phosphopeptides**. *Anal Chem* 1999, **71**:2883-2892.

61. • Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM: **Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae***. *Nat Biotechnol* 2002, **20**:301-305.

This paper describes the improvement of specificity of IMAC-based phospho-peptides enrichment by methyl-esterification.

62. Haydon CE, Eyers PA, Aveline-Wolf LD, Resing KA, Maller JL, Ahn NG: **Identification of novel phosphorylation sites on *Xenopus laevis* Aurora A and analysis of phosphopeptide enrichment by immobilized metal-affinity chromatography**. *Mol Cell Proteomics* 2003, **2**:1055-1067.

63. Kalume DE, Molina H, Pandey A: **Tackling the phosphoproteome: tools and strategies**. *Curr Opin Chem Biol* 2003, **7**:64-69.

64. Zhou H, Watts JD, Aebersold R: **A systematic approach to the analysis of protein phosphorylation**. *Nat Biotechnol* 2001, **19**:375-378.

65. Oda Y, Nagasu T, Chait BT: **Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome**. *Nat Biotechnol* 2001, **19**:379-382.

66. Goshe MB, Conrads TP, Panisko EA, Angell NH, Veenstra TD, Smith RD: **Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses**. *Anal Chem* 2001, **73**:2578-2586.

67. Weckwerth W, Willmitzer L, Fiehn O: **Comparative quantification and identification of phosphoproteins using stable isotope labeling and liquid chromatography/mass spectrometry**. *Rapid Commun Mass Spectrom* 2000, **14**:1677-1681.

68. •• Knight ZA, Schilling B, Row RH, Kenski DM, Gibson BW, Shokat KM: **Phosphospecific proteolysis for mapping sites of protein phosphorylation**. *Nat Biotechnol* 2003, **21**:1047-1054.

This is an excellent example of method development on mapping serine/threonine phosphorylated sites by proteolysis specific to the phosphorylated sites.

69. Marotti LA Jr, Newitt R, Wang Y, Aebersold R, Dohlman HG: **Direct identification of a G protein ubiquitination site by mass spectrometry**. *Biochemistry* 2002, **41**:5067-5074.

70. •• 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**:921-926.

This paper provides the first demonstration of large-scale proteomics analysis on protein ubiquitination.

71. Schwartz DC, Hochstrasser M: **A superfamily of protein tags: ubiquitin, SUMO and related modifiers**. *Trends Biochem Sci* 2003, **28**:321-328.

72. Liu Y, Patricelli MP, Cravatt BF: **Activity-based protein profiling: the serine hydrolases**. *Proc Natl Acad Sci USA* 1999, **96**:14694-14699.

73. Greenbaum D, Medzihradszky KF, Burlingame A, Bogyo M: **Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools**. *Chem Biol* 2000, **7**:569-581.

74. • Lo LC, Pang TL, Kuo CH, Chiang YL, Wang HY, Lin JJ: **Design and synthesis of class-selective activity probes for protein tyrosine phosphatases**. *J Proteome Res* 2002, **1**:35-40.

An activity-based protein probe for phosphatase is discussed.

75. •• Leung D, Hardouin C, Boger DL, Cravatt BF: **Discovering potent and selective reversible inhibitors of enzymes in complex proteomes**. *Nat Biotechnol* 2003, **21**:687-691.

This paper discusses competitive profiling of serine hydrolyse inhibitors using chemical probes and provides a potential method for screening enzyme inhibitors.

76. •• Adam GC, Sorensen EJ, Cravatt BF: **Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype**. *Nat Biotechnol* 2002, **20**:805-809.

This paper describes a study using a library with a sulfonate ester chemotype to screen complex proteomes for activity-dependent protein activities and provides a useful method for future drug screen.