Proteomic analysis of post-translational modifications

Matthias Mann[1]* and Ole N. Jensen[2]*

Post-translational modifications modulate the activity of most eukaryote proteins. Analysis of these modifications presents formidable challenges but their determination generates indispensable insight into biological function. Strategies developed to characterize individual proteins are now systematically applied to protein populations. The combination of function- or structure-based purification of modified 'subproteomes', such as phosphorylated proteins or modified membrane proteins, with mass spectrometry is proving particularly successful. To map modification sites in molecular detail, novel mass spectrometric peptide sequencing and analysis technologies hold tremendous potential. Finally, stable isotope labeling strategies in combination with mass spectrometry have been applied successfully to study the dynamics of modifications.

Post-translational modifications (PTMs) are covalent processing events that change the properties of a protein by proteolytic cleavage or by addition of a modifying group to one or more amino acids. Far from being mere "decorations", PTMs of a protein can determine its activity state, localization, turnover, and interactions with other proteins. In signaling, for example, kinase cascades are turned on and off by the reversible addition and removal of phosphate groups[1], and in the cell cycle ubiquitination marks cyclins for destruction at defined time points[2]. Table 1 summarizes the features of a few important PTMs.

Despite the great importance of PTMs for biological function, their study on a large scale has been hampered by a lack of suitable methods, and many key modifications have only been discovered late in the elucidation of various biological processes. As a result, we probably still do not realize the full extent and functional importance of protein modifications in the workings of the cell.

Many PTMs have been discovered serendipitously during studies of individual proteins with the help of standard molecular techniques, such as deletion of the amino acids bearing the modification. Direct analysis of modifications requires isolation of the correctly processed protein in a sufficiently large amount for biochemical study—a feat that has only been possible in selected cases.

The introduction, first of improved preparation and derivatization methods for Edman degradation, and then of mass spectrometry (MS) methods, has continuously increased the sensitivity for detection of PTMs. MS measures mass-to-charge ratio (m/z), yielding the molecular weight and the fragmentation pattern of peptides derived from proteins. It therefore represents a general method for all modifications that change the molecular weight. For this reason, MS has long been used to map the complete primary structure of individual proteins (e.g., see refs. 3–5).

The new field of proteomics is concerned with structural and functional properties of large sets of proteins[6]. Proteomics has been very successful in identifying proteins in complexes and organelles, and hundreds of proteins can now be analyzed in a single experiment. However, these investigations have typically mapped only a few peptides per protein; this is sufficient for identification but not for complete primary-structure determination.

In this review, we explore current techniques for mapping post-translational modifications and examine their applicability at the proteomic scale, an endeavor sometimes called "modification-specific proteomics"[7]. Several recently developed approaches, especially in phosphorylation analysis (for recent reviews, see refs. 8–11), are beginning to yield results for proteome-wide PTM analysis. The complete characterization of the primary structure of large populations of proteins, however, will remain a challenging and fruitful area for proteomics for many years to come.

Isolation of modified proteins

Modification analysis is usually done by comparison of experimental data to a known amino acid sequence. Therefore, the first step is identification of the protein to be studied, which can be done at very high sensitivity by antibody recognition (western blotting) or by MS techniques.

A central consideration in the characterization of modifications is the need for as large an amount of the protein as possible. Protein modifications are typically not homogeneous, and a single gene can give rise to a bewildering number of gene products as a result of alternative splicing and the combination of different modifications. The amount of protein in a single modification state can thus be a very small fraction of the total amount of the gene product. Furthermore, as explained later, the complete characterization of the primary structure of a protein requires much more material than mere identification by MS sequencing of a few peptides.

For these reasons, and because therapeutic protein products of the biotechnology industry require thorough characterization, much work has been done on recombinantly expressed proteins or recombinantly expressed proteins that are modified *in vitro*. In addition, although expression systems, such as baculovirus, are thought to produce modification patterns similar to those of mammalian cells, there are often significant differences.

[1]*Center for Experimental BioInformatics (CEBI) and* [2]*Protein Research Group, Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, Odense M, DK-5230 Denmark. *Corresponding authors (mann@bmb.sdu.dk) and (jenseno@bmb.sdu.dk).*

Table 1. Some common and important post-translational modifications

PTM type	ΔMass[a] (Da)	Stability[b]	Function and notes
Phosphorylation			Reversible, activation/inactivation of enzyme activity, modulation of molecular interactions, signaling
pTyr	+80	+++	
pSer, pThr	+80	+/++	
Acetylation	+42	+++	Protein stability, protection of N terminus. Regulation of protein–DNA interactions (histones)
Methylation	+14	+++	Regulation of gene expression
Acylation, fatty acid modification			Cellular localization and targeting signals, membrane tethering, mediator of protein–protein interactions
Farnesyl	+204	+++	
Myristoyl	+210	+++	
Palmitoyl	+238	+/++	
etc.			
Glycosylation			
N-linked	>800	+/++	Excreted proteins, cell–cell recognition/signaling
O-linked	203, >800	+/++	*O*-GlcNAc, reversible, regulatory functions
GPI anchor	>1,000	++	Glycosylphosphatidylinositol (GPI) anchor. Membrane tethering of enzymes and receptors, mainly to outer leaflet of plasma membrane
Hydroxyproline	+16	+++	Protein stability and protein–ligand interactions
Sulfation (sTyr)	+80	+	Modulator of protein–protein and receptor–ligand interactions
Disulfide bond formation	–2	++	Intra- and intermolecular crosslink, protein stability
Deamidation	+1	+++	Possible regulator of protein–ligand and protein–protein interactions, also a common chemical artifact
Pyroglutamic acid	–17	+++	Protein stability, blocked N terminus
Ubiquitination	>1,000	+/++	Destruction signal. After tryptic digestion, ubiquitination site is modified with the Gly-Gly dipeptide
Nitration of tyrosine	+45	+/++	Oxidative damage during inflammation

[a]A more comprehensive list of PTM Δmass values can be found at: http://www.abrf.org/index.cfm/dm.home
[b]Stability: + labile in tandem mass spectrometry, ++ moderately stable; +++ stable.

Two-dimensional gel electrophoresis separates protein populations on the basis of charge and molecular weight (for recent reviews, see refs. 12–14). It often has sufficient resolution to separate the modification states of a protein directly. For example, phosphorylation changes the protein charge and is often indicated by a horizontal trail of protein spots on the two-dimensional gel. If there is sufficient material, each of these spots can be analyzed by the methods described below, to define the modified amino acid residues. In one study, N- and C-terminal processing of enolase was studied by a combination of two-dimensional electrophoresis and MS, defining >10 different forms[15].

The staining intensity of two-dimensional gel electrophoresis spots roughly reflects the protein amount, especially for similar proteins; therefore, this method also provides information about the relative proportion of the various modified states. Two-dimensional gels can also be used directly with cell or tissue lysates; however, it is often advantageous to prefractionate to reduce the complexity of the mixture and increase the protein amount that can be loaded onto the gel.

To study the modifications of a single protein, chromatographic purifications, antibody precipitations, or both are the usual methods used to isolate sufficient amounts. Modern analysis methods can tolerate protein contamination much better than earlier methods; therefore, the total amount of recovered protein is more important than absolute purity. Often, sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) will be the final preparation step, and researchers should attempt to isolate at least Coomassie-stainable amounts (several picomoles or ~1 μg) to increase the chance of detecting and characterizing modifications in proteins.

PTM mapping of a purified protein

Once a protein has been isolated, a variety of techniques can be used to determine the modified amino acids. In some cases, the precise molecular weight of the intact protein can be established by MS, especially if the protein is not too heterogeneous, its mass is less than about 100 kDa, and it is in a buffer that is compatible with MS. Once the masses of the nonmodified and modified amino acid residues add up to the measured intact molecular weight, the protein is completely characterized.

Amino-terminal protein sequencing by the classical technique of Edman degradation is still the method of choice to determine proteolytic processing. Carboxy-terminal processing can also be determined by amino acid sequencing, albeit at a much lower sensitivity.

Detailed characterization of modification happens after enzymatic or chemical degradation of the protein (see Fig. 1). The resulting peptides are usually separated by high-performance liquid chro-

matography (HPLC). In Edman degradation, collected peptide fractions are applied to the sequencer and their amino acid sequence determined. Modified amino acids become apparent because of their absence or retention-time shift in the corresponding sequencing cycle. If the mass of the intact peptide has been determined, then the nature of the modification can be more confidently assessed. One method to determine phosphorylation uses a combination of MS and Edman degradation: proteins are labeled with ^{32}P, and the cycle in which the radiolabeled amino acid is released is recorded. Together with the mass of the peptide, this often allows determination of the site of phosphorylation[16,17]. Thin-layer chromatography in combination with ^{32}P labeling is still widely used for phosphopeptide mapping and phosphoamino acid determination.

Figure 1 shows the general strategy of PTM mapping for a single protein. The protein is enzymatically digested into a collection of peptides that are more readily characterized than the intact protein. The peptide mixture can be analyzed directly by matrix-assisted laser desorption/ionization (MALDI) peptide mapping or nanoelectrospray MS. Even though the starting amount of peptides from a given protein is the same, a complex and incompletely understood series of factors result in very different signal intensities for each of these peptides. Thus, some of the peptides will be clearly visible and some will not be apparent at all. This is particularly the case for complex mixtures and for modified peptides that often do not ionize as well as nonmodified peptides.

As a result, peptides are typically separated by reversed-phase chromatography, with subsequent fraction collection and analysis, or in combination with direct "online" electrospray mass spectrometry (LC/MS). Even this combination does not necessarily result in measurement of peptides covering the complete protein ("100% sequence coverage"); for example, very short peptides of only a few amino acids or very long peptides of >30 amino acids are often not detected.

The latter problem can be alleviated by means of a second experiment in which the protein is digested by another enzyme with a different specificity. The first choice of enzyme is usually trypsin, which cleaves C-terminal to arginine and lysine, therefore resulting in peptides with basic amino acids at the C terminus, which are most easily sequenced by MS. A second enzyme can then be endoproteinase Asp-N or Glu-C, which are also very specific. Note, however, that there is still no assurance that every part of the protein is covered. Even if this were the case, a minor, modified form of a peptide may still escape detection (Fig. 1). This problem can be addressed by targeted methods, as described later.

The experimentally determined peptide masses are matched against the list of peptide masses expected from the protein sequence. The nonmatching masses can arise as a result of contaminating proteins, unexpected cleavage, or common chemical modifications (e.g., methionine oxidation). The remaining masses are inspected for mass differences compared with the expected peptides that correspond to a modification (see Table 1 for some of these "Δ mass" values).

Often, the peptide mass pattern will already hint at the nature of the modification, such as multiple mass differences of 162 Da for glycosylation and the presence of a "satellite mass" of –98 Da in the case of phosphoserine and phosphothreonine because of the elimination of phosphoric acid.

The mass of the modified peptide is usually not sufficient to determine the nature of the modification, and peptides are fragmented by MS to identify the peptide and localize its modification (see Fig. 2). In these "tandem mass spectrometry" (MS/MS) experiments, peptide ions are collided with inert gas, leading to fragmentation, usually at the peptide bonds. Some modified amino acid residues remain intact during this process. In this case, the fragmentation pattern is

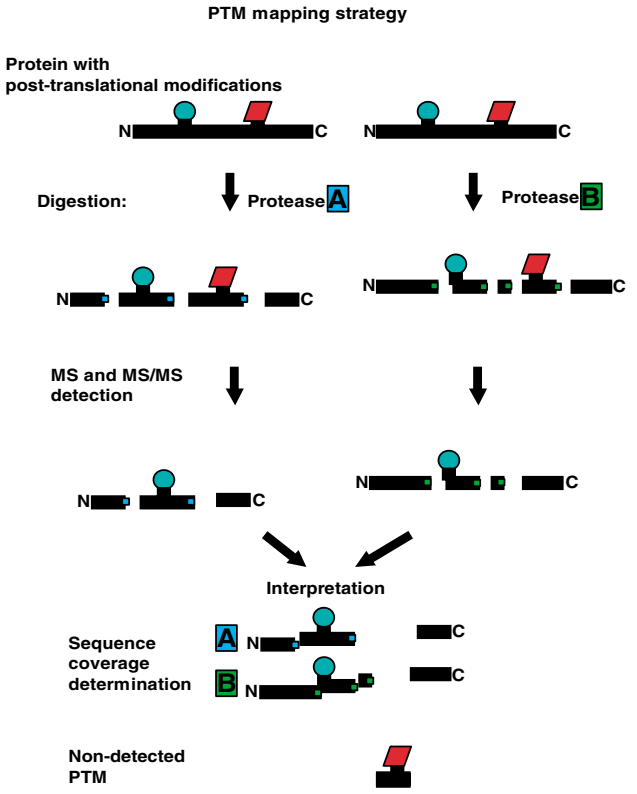

Figure 1. Strategy for mapping post-translational modifications. The protein is shown as a line, and modifications are indicated by the symbols. The two columns show enzymatic digestion by two different enzymes to cover as much as possible of the protein sequence with peptides in the preferred mass range for MS analysis (500–3,000 Da).

similar to the unmodified peptide with the difference that the location of the modified amino acid is revealed by its mass increment. Thus, ideally the peptide sequence and the mass and location of the modification can be determined. In practice, the fragmentation pattern may or may not allow exact localization of the modification, depending on the completeness of the fragmentation pattern.

If the modification is labile, then it will be lost before the peptide itself fragments. In this case, the peptide can still be sequenced and identified, but only the mass increment (not the location of the modification) is determined.

Examples of stable modifications are acetylation (+42 Da)—which is found on the N termini of many proteins or on specific lysine residues—and arginine methylation (+14 Da). Examples of labile modifications are O-linked *N*-acetylglucosamine (GlcNAc; +203 Da) and sulfation (+80 Da). The phosphogroup (+80 Da) can be stable (e.g., in the case of phosphotyrosine) or relatively labile (e.g., in the case of phosphothreonine and especially phosphoserine; Table 1).

Modifying groups that are easily lost from the peptide can themselves be used as "reporter groups" to detect the presence of the modified peptide in several different ways: In "in-source fragmentation", excess energy in the ionization or ion-sampling process leads to the characteristic presence of the reporter ion in mass spectra. Subsequent sequencing of the peptide peaks can then identify the modified peptide.

Conversely, in the "neutral loss" technique, mild collisions in the collision cell between the two sections of a tandem mass spectrometer lead to loss of the modifying group[18,19]. The second mass analyzer is set at a mass offset corresponding to the mass-to-charge ratio

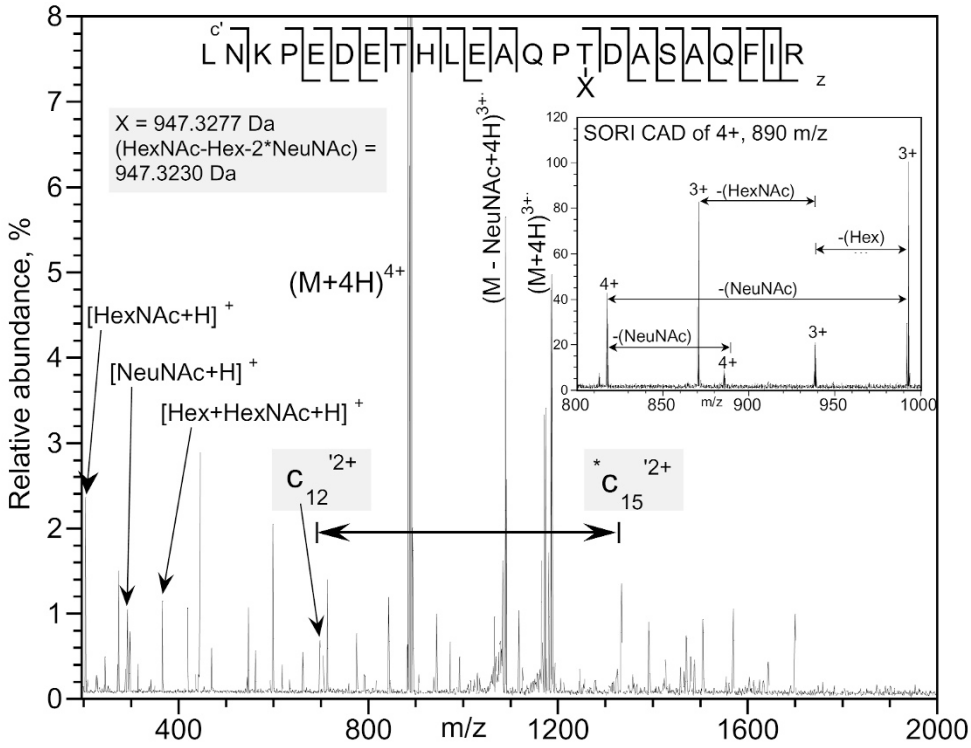

Figure 2. Determination of a glycosylation site by FTMS and ECD. A glycosylated peptide from the milk protein PP3 was ionized by electrospray and analyzed in an FTMS mass spectrometer. Upon fragmentation in the conventional way by collisions, the only products are sugar losses for this labile modification (inset), yielding information about the nature of the modification but not the peptide sequence or the location of the modification. Upon fragmentation by ECD, the N- and C-terminal fragments denoted in the sequence were observed. The sugar modification was localized to the threonine by the c_{12} and c_{15} fragments indicated in the spectrum. The composition of the sugar modification denoted X in the sequence was determined to be $HexNAc_1Hex_1NeuNAc_2$. In total, five phosphorylations, three O- and one N-glycosylation, were determined in the protein. Adapted and reprinted with permission from ref. 36 (©2003 American Chemical Society).

(*m/z*) of the modification. A signal can only reach the detector if the peptides were modified and the mass changed by the expected amount during collision.

Finally, in a related technique named "precursor-ion scans", the second mass analyzer is set at the mass of the reporter ion while the first analyzer sweeps through the mass range[20,21]. Only peptides that generate the reporter ion are registered in the final spectrum, and modified peptides can specifically be detected in an excess of background ions. A high-resolution variant of this technique has successfully been used to map phosphotyrosine modifications in the epidermal growth factor (EGF) and fibroblast growth factor pathways[22–24].

Normally, peptides are identified in databases by their tandem mass spectrum using pattern matching with a variety of different algorithms, such as Peptide Sequence Tags[25], Sequest[26], or Mascot[27]. The last two were initially developed for unmodified peptides but have been extended to modified peptides[28,29]. This is typically done by taking more than one possible amino acid molecular weight into account, depending on the modification considered. For example, it can be specified that tyrosine can be unmodified or that it can have the molecular weight of phosphotyrosine (i.e., the molecular weight can be 163 Da or 243 Da), which will change the calculated peptide molecular weight and that of all fragments containing the modified tyrosine. It should be noted, however, that identifications of modified peptides are often less certain than those of unmodified peptides, because there are, in effect, more peptides against which to match.

Especially when allowing a number of different modifications, the "combinatorial explosion" of different peptide forms can lead to problems. Moreover, tandem mass spectra of peptides with labile modifications are often less informative, because the majority of ions undergo only elimination of the modification. Nevertheless, in the case of a single protein discussed here, these difficulties can often be overcome by detailed validation of the tandem mass spectrum against possible modified tryptic peptides predicted from the sequence.

A recommended practice would be to purify at least 1 μg (Coomassie-stainable amount) of protein, enzymatically digest it using first trypsin and then in a second experiment another sequence-specific protease, and analyze these digests with liquid chromatography tandem mass spectrometry (LC MS/MS) and a database searching program (Fig. 1). These experiments should yield most modifications, except those of very low stoichiometry, or those in areas not "covered" by either of the proteases. To investigate any remaining modifications, or if only specific modifications are of interest, some of the more specialized techniques mentioned already should be applied.

Technologies now in development may substantially accelerate the laborious task of PTM analysis. Fourier-transform mass spectrometry (FTMS) instruments are ion traps with extremely high mass resolution and potentially very high sensitivity[30,31]. These features allow measurement of the precise intact molecular weight of proteins, as already mentioned, and facilitate the analysis of exceedingly complex peptide mixtures.

Additionally, FTMS can be coupled with electron capture dissociation (ECD), a relatively new technology with great potential for PTM mapping[32–36]. In this technique, dissociation is induced by electron recombination with the protons of the multiply charged peptide or protein. Intriguingly, even labile modifications remain

intact during this process, allowing unambiguous assignment in the peptide sequence (Fig. 2). In a technique called "top-down" protein sequencing, the intact protein is fragmented, theoretically allowing complete mapping of modifications, all inside the mass spectrometer[34,37].

PTM mapping of protein populations

Although the methods just described are very powerful for the characterization of individual, purified proteins and have helped elucidate numerous biological mechanisms, the real promise of proteomics is to assess systematically the modifications of large numbers of proteins[7]. At this stage, there are basically four different experimental approaches to this problem.

Two-dimensional gel-based PTM analysis. As mentioned earlier, protein modification states are often separated by two-dimensional gel electrophoresis. In one approach, modified proteins are specifically visualized on gels or on membranes. For example, phosphorylated proteins can be recognized by anti-phosphoamino acid antibodies. These spots can then be excised and identified by MS in the standard manner[38]. Alternatively, it might be possible to visualize modified "subproteomes" by metabolic labeling, again followed by mass spectrometric spot identification (N.V. Bykova and O.N. Jensen, unpublished data). Finally, protein populations can be run on two-dimensional gels before and after enzymatic removal of the modifying group[39]. "Disappearing" protein spots are then an indication of the modification in question.

Protein spots can also be mapped in detail to determine the exact site and nature of the modification, using the methods described earlier for single-protein analysis. Note, however, that the usually very small protein amounts in two-dimensional gel spots may result in low sequence coverage and therefore a low probability of finding the modified peptides. This problem can sometimes be alleviated by affinity enrichment of modified peptides[40], as further described later.

Two-dimensional gels visualize only a subset of the proteome, especially if only one condition is used for the second dimension; moreover, not all modification states will be separated from each other. Furthermore, more abundant proteins may co-migrate with the modified protein of interest. In this case, it may not be clear which of the identified proteins is modified, if the actual modified peptide has not been detected.

Affinity-based enrichment of modified proteins. A second approach for PTM detection combines established biochemical, genetic, and immunological methods for enrichment of modified-protein populations with the recently developed MS techniques for protein mixture analysis. This strategy is particularly attractive because the enrichment step is often a single experiment (e.g., an immunoprecipitation), and the subsequent identification of the protein mixture is usually reduced to a single LC MS/MS experiment as well.

Glycosylphosphatidylinositol (GPI) anchors tether proteins to the extracellular leaflet of the plasma membrane, where they act as enzymes, receptors, and adapters. Work in our laboratory (F. Elortza and O.N. Jensen, unpublished data) has shown that this protein class (of obvious pharmacological interest) can be selectively released by enzymatic cleavage of the anchor and extracted from membrane preparations. Protein analysis by LC MS/MS resulted in dozens of protein identifications, covering a large percentage of the predicted GPI-anchored proteome.

The phosphoproteome has also been explored with this strategy. For example, cells were stimulated with EGF, and proteins that had just been tyrosine-phosphorylated were immunoprecipitated with anti-phosphotyrosine antibody[41]. The phosphoamino acid antibody–based enrichment strategy has also been used for phosphoserine and threonine[42].

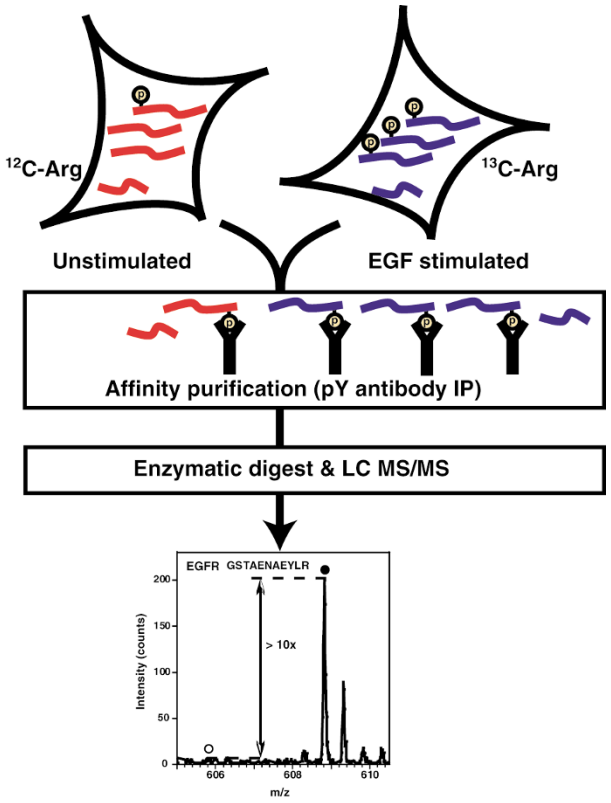

Figure 3. Analysis of the phosphoproteome. In the example, one population of cells is metabolically labeled with [^{13}C]arginine, which results in a mass shift with respect to the "normal" arginine-labeled cell population for all arginine-containing peptides[58]. The [^{13}C]arginine-labeled cells are stimulated, resulting in phosphotyrosine phosphorylation of a population of proteins. Cell lysates are combined, and the phosphotyrosine-phosphorylated "subproteome" is affinity-purified using an antibody specific for phosphotyrosine. After enzymatic digestion and MS analysis, peptides that are phosphorylated in the stimulated and nonstimulated population can be quantified by the ratios of their peak heights. As an example, the peptide GSTAENAEYLR (one-letter amino acid code) of the EGF receptor is found in the labeled form, indicating that the EGF receptor is phosphorylated upon EGF stimulation (ref. 59; B. Blagoev, S.-E. Ong, and M. Mann, unpublished data).

Another modification of great interest, the enzymatic attachment of ubiquitin to cellular proteins that marks them for destruction, is also under investigation. In an elegant experiment, yeast ubiquitin was replaced by a histidine-tagged version, allowing selective purification and identification of the ubiquitinated proteome (ref. 43; S. P. Gygi, personal communication). In a subset of the proteins, the site of ubiquitination could also be defined.

As these examples show, the combination of selective enrichment of modified proteins with MS mixture analysis can be very powerful. The critical step is the development of the enrichment protocol. Subsequently, the proteins have only to be identified, thus avoiding the difficulties of detailed modification mapping mentioned earlier.

Identification of PTMs in complex peptide mixtures. Recent technological developments have made it increasingly feasible to directly analyze very complex peptide mixtures by LC MS/MS, and a single chromatographic run can result in the identification of hundreds of peptides and proteins.

Unfortunately, all but the most abundant proteins in a mixture are "covered" only by a few sequenced peptides. This is because peptide sequencing takes ~1 second per peptide, limiting the num-

ber of sequencing events. Furthermore, mass spectrometers have a finite dynamic range, and peptides of very low abundance may not be picked for sequencing in the presence of peptides that are very abundant.

Nevertheless, experiments of this type often fragment thousands of peptides, so the chance of finding modified peptides is also substantial. Furthermore, as in the case of single-protein analysis, the proteins can be digested with several proteases, improving sequence coverage. Exemplifying this approach, a study of eye lens crystalline proteins reported a large number of modified peptides[44].

Currently, large-scale protein identification projects, involving the fragmentation and identification of many proteins, can "harvest" protein modifications by a simple two-step process. In a first step, proteins are identified permitting only unmodified peptides in the database searches. In the second step, this "database' of identified proteins is interrogated further with hitherto-unassigned fragmentation spectra, allowing the possibility of protein modification. In our laboratories, this is routinely done for protein complexes involved in signaling processes. This strategy does not compromise the certainty of identification of proteins present, and as long as this list of proteins is not too large, modified peptides can be found in this very restricted set of proteins with high confidence.

The chromatographic and mass spectrometric techniques for analyzing protein mixtures are continuously being improved; therefore, the chances of finding specific and functionally relevant modifications in complex mixtures, though slim at this time, are also continuously improving. We expect that only minor changes in the experimental protocol (e.g., the use of two proteolytic enzymes instead of one) will be needed to make possible the scanning of PTMs in small protein complexes, such as the ones analyzed in global protein–protein interaction studies[45,46].

Derivatization and affinity-based methods. As mentioned earlier, peptide mixtures derived from complex protein mixtures are very difficult to analyze comprehensively. If one is interested in specific modifications, the peptide complexity can be reduced by affinity methods. For example, phosphopeptides can be captured selectively through their negatively charged phosphogroup on immobilized-metal affinity (IMAC) columns[40,47,48]. Recently, this technique has been made much more specific by esterifying and thereby neutralizing the negatively charged amino acid residues before the IMAC step[49], allowing identification of hundreds of phosphopeptides in yeast cell lysates. The method has also been used in combination with phosphotyrosine protein affinity purification[50].

Chemical derivatization of the modifying group potentially allows attachment of a "hook" for affinity purification. For example, the phosphate group can be converted to an affinity tag by a β-elimination–Michael addition reaction or by phosphoamidate chemistry[51–53]. It should be noted, however, that only very simple and extremely efficient chemical derivatization steps are compatible with proteomics. If any heterogeneity is introduced by the chemical reaction (e.g., as a result of <100% conversion efficiency or side reactions), the peptide samples become even more complex and it is then only possible to analyze modifications of the most abundant proteins.

Conclusions and perspectives

Classic techniques and mass spectrometry are both widely used in the analysis of individual, purified proteins. To characterize any single protein in complete detail is still a substantial research project, which by proteomic standards requires large amounts of protein and a combination of different techniques.

For the characterization of protein populations, the innovative mass spectrometric methods offer great promise. As examples from phosphorylation analysis and others show, affinity enrichment of modified protein populations combined with mass spectrometric identification of the "subproteome" so defined has already achieved success.

Challenges for the future include obtaining much higher sequence coverage of proteins in complex mixtures. This is a precondition of any comprehensive study of protein modifications and may be achieved by higher digestion efficiencies, higher sensitivity of peptide analysis, and the ability to sequence many more peptides in a single experiment. Selective sequencing of only the modified peptides may be achievable by building databases of previously observed peptides, which are then excluded from sequencing.

In terms of software and hardware, database search algorithms are not optimized for the detection of modified peptides. Further development in this field would benefit proteomics tremendously. Current instrument developments of interest for PTM analysis include MALDI MS/MS[54] and the combination of ECD peptide fragmentation with new generations of high-sensitivity FTMS instruments mentioned earlier. ECD with FTMS in particular offers the hope of characterizing labile modifications with a sensitivity and speed approaching standard mass-spectrometric peptide sequencing.

Given the technical difficulties involved, much of PTM analysis has concentrated on identifying modifications, rather than on quantifying them. However, biological effects are often due to changes in the level of modification, and therefore there is intense interest in the quantitative study of modifications. Stable isotope labeling in various forms is now being used to quantify protein levels on a large scale[55]. Similar techniques can also be applied to modifications (Fig. 3). In one study, stable isotope labeling was used for relative quantification of phosphorylation sites in yeast[56], and in another study synthesized stable isotope analogs of phosphopeptides provided absolute quantification[57].

Bioinformatic methods can predict modification sites *in silico*. Although they are extremely promising and important, however, these methods are no substitute for direct experimental data. Also, because most modifications of interest are regulatory and reversible, they cannot be predicted from the sequence alone. As proteomics defines more and more *in vivo* modification sites, prediction algorithms can be tuned and improved with this information. Conversely, bioinformatic prediction tools should be closely integrated into the interpretation of proteomic PTM experiments.

Proteomics is now sufficiently developed that thought should be given to databases of protein modifications similar to the genomic and protein sequence databases in existence today. Such databases should contain not only the nature and sites of modifications, but also the primary data, such that future proteome experiments can be interpreted in reference to these data, instead of the more difficult interpretation of results *de novo*.

Once PTM analysis can routinely be done at the proteomic level, the involvement of PTMs in disease can be studied much more systematically than has thus far been possible. Although many examples of PTMs in disease are known, it is very likely that these examples are just the "tip of the iceberg". Proteomic PTM analysis will thus contribute to our understanding of disease etiology and deliver many new targets for research against diseases.

Acknowledgments

We thank R. Zubarev, Uppsala University, for providing Figure 2. Work in the authors' laboratories is supported by a generous grant of the Danish National Research Foundation (M.M.) and the Danish Natural Sciences Research Council (O.N.J.) and infrastructure provided by the Danish Biotechnology Instrument Center(M.M. and O.N.J.).

Received 27 January 2003; accepted 10 February 2003

1. Cohen, P. The regulation of protein function by multisite phosphorylation—a 25 year update. *Trends Biochem. Sci.* **25**, 596–601 (2000).
2. Tyers, M. & Jorgensen, P. Proteolysis and the cell cycle: with this RING I do thee destroy. *Curr. Opin. Genet. Dev.* **10**, 54–64 (2000).
3. Carr, S.A. *et al.* Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor by mass spectrometry. *J. Biol. Chem.* **264**, 21286–21295 (1989).
4. Ling, V. *et al.* Characterization of the tryptic map of recombinant DNA-derived tissue plasminogen activator by high-performance liquid chromatography–electrospray ionization mass spectrometry. *Anal. Chem.* **63**, 2909–2915 (1991).
5. Roepstorff, P. Mass spectrometry in protein studies from genome to function. *Curr. Opin. Biotechnol.* **8**, 6–13 (1997).
6. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature,* in press (2003).
7. Jensen, O.N. Modification-specific proteomics: strategies for systematic studies of post-translationally modified proteins. in *Proteomics: A Trends Guide* (eds. Blackstock, W. & Mann, M.) 36–42 (Elsevier Science, London; 2000).
8. McLachlin, D.T. & Chait, B.T. Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr. Opin. Chem. Biol.* **5**, 591–602 (2001).
9. Sickmann, A. & Meyer, H.E. Phosphoamino acid analysis. *Proteomics* **1**, 200–206 (2001).
10. Mann, M. *et al.* Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol.* **20**, 261–268 (2002).
11. Loughrey Chen, S. *et al.* Mass spectrometry-based methods for phosphorylation site mapping of hyperphosphorylated proteins applied to Net1, a regulator of exit from mitosis in yeast. *Mol. Cell Proteomics* **1**, 186–196 (2002).
12. Gorg, A. *et al.* The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**, 1037–1053 (2000).
13. Rabilloud, T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* **2**, 3–10 (2002).
14. Fey, S.J. & Larsen, P.M. 2D or not 2D. Two-dimensional gel electrophoresis. *Curr. Opin. Chem. Biol.* **5**, 26–33 (2001).
15. Larsen, M.R., Larsen, P.M., Fey, S.J. & Roepstorff, P. Characterization of differently processed forms of enolase 2 from *Saccharomyces cerevisiae* by two-dimensional gel electrophoresis and mass spectrometry. *Electrophoresis* **22**, 566–575 (2001).
16. Knebel, A., Morrice, N. & Cohen, P. A novel method to identify protein kinase substrates: eEF2 kinase is phosphorylated and inhibited by SAPK4/p38δ. *EMBO J.* **20**, 4360–4369 (2001).
17. MacDonald, J.A., Mackey, A.J., Pearson, W.R. & Haystead, T.A. A strategy for the rapid identification of phosphorylation sites in the phosphoproteome. *Mol. Cell Proteomics* **1**, 314–322 (2002).
18. Covey, T.R., Shushan, B.I., Bonner, R., Schroder, W. & Hucho, F. LC/MS and LC/MS/MS screening for the sites of post-translational modifications in proteins. in *Methods in Protein Sequence Analysis* (eds Jörnvall, H., Höög, J.O. & Gustavsson, A.M.). 249–256 (Birkhäuser Verlag, Basel, 1991).
19. Bateman, R.H. *et al.* A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *J. Am. Soc. Mass Spectrom.* **13**, 792–803 (2002).
20. Wilm, M., Neubauer, G. & Mann, M. Parent ion scans of unseparated peptide mixtures. *Anal. Chem.* **68**, 527–533 (1996).
21. Carr, S.A., Huddleston, M.J. & Annan, R.S. Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry. *Anal. Biochem.* **239**, 180–192 (1996).
22. Steen, H., Kuster, B., Fernandez, M., Pandey, A. & Mann, M. Detection of tyrosine-phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. *Anal. Chem.* **73**, 1440–1448 (2001).
23. Steen, H., Kuster, B., Fernandez, M., Pandey, A. & Mann, M. Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J. Biol. Chem.* **277**, 1031–1039 (2002).
24. Hinsby, A.M., Olsen, J.V., Bennett, K.L. & Mann, M. Signaling initiated by overexpression of the fibroblast growth factor receptor-1 investigated by mass spectrometry. *Mol. Cell. Proteomics,* in press (2003).
25. Mann, M. & Wilm, M.S. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**, 4390–4399 (1994).
26. Eng, J.K., McCormack, A.L. & Yates, J.R. I. An approach to correlate MS/MS data to amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
27. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
28. MacCoss, M.J., Wu, C.C. & Yates, J.R. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal. Chem.* **74**, 5593–5599 (2002).
29. Creasy, D.M. & Cottrell, J.S. Error-tolerant searching of uninterpreted tandem mass spectrometry data. *Proteomics* **2**, 1426–1434 (2002).
30. Marshall, A.G., Hendrickson, C.L. & Jackson, G.S. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **17**, 1–35 (1998).
31. Martin, S.E., Shabanowitz, J., Hunt, D.F. & Marto, J.A. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **72**, 4266–4274 (2000).
32. Zubarev, R.A. *et al.* Electron-capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* **72**, 563–573 (2000).
33. Stensballe, A., Jensen, O.N., Olsen, J.V., Haselmann, K.F. & Zubarev, R.A. Electron-capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun. Mass Spectrom.* **14**, 1793–1800 (2000).
34. Shi, S.D. *et al.* Phosphopeptide/phosphoprotein mapping by electron-capture dissociation mass spectrometry. *Anal. Chem.* **73**, 19–22 (2001).
35. Kelleher, N.L. *et al.* Localization of labile posttranslational modifications by electron-capture dissociation: the case of γ-carboxyglutamic acid. *Anal. Chem.* **71**, 4250–4253 (1999).
36. Kjeldsen, F., Haselmann, K.F., Budnik, B.A., Soerensen, E.S. & Zubarev, R.A. Complete characterization of post-translational modification sites in the bovine milk protein PP3 by tandem mass spectrometry with electron-capture dissociation as the last stage. *Anal. Chem.*, in press (2003).
37. Sze, S.K., Ge, Y., Oh, H. & McLafferty, F.W. Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl. Acad. Sci. USA* **99**, 1774–1779 (2002).
38. Soskic, V., Gorlach, M., Poznanovic, S., Boehmer, F.D. & Godovac-Zimmermann, J. Functional proteomics analysis of signal transduction pathways of the platelet-derived growth factor-β receptor. *Biochemistry* **38**, 1757–1764 (1999).
39. Yamagata, A. *et al.* Mapping of phosphorylated proteins on two-dimensional polyacrylamide gels using protein phosphatase. *Proteomics* **2**, 1267–1276 (2002).
40. Stensballe, A., Andersen, S. & Jensen, O.N. Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics* **1**, 207–222 (2001).
41 Pandey, A. *et al.* Analysis of receptor signaling pathways by mass spectrometry: identification of Vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc. Natl. Acad. Sci. USA* **97**, 179–184 (2000).
42. Gronborg, M. *et al.* A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol. Cell. Proteomics* **1**, 517–527 (2002).
43. Peng, J. & Gygi, S.P. Proteomics: the move to mixtures. *J. Mass Spectrom.* **36**, 1083–1091 (2001).
44. MacCoss, M.J. *et al.* Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. USA* **99**, 7900–7905 (2002).
45. Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
46. Gavin, A.C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
47. Andersson, L. & Porath, J. Isolation of phosphoproteins by immobilized metal (Fe^{3+}) affinity chromatography. *Anal. Biochem.* **154**, 250–254 (1986).
48. Posewitz, M.C. & Tempst, P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* **71**, 2883–2892 (1999).
49. Ficarro, S.B. *et al.* Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **20**, 301–305 (2002).
50. Salomon, A.R. *et al.* Profiling of tyrosine-phosphorylation pathways in human cells using mass spectrometry. *Proc. Natl. Acad. Sci. USA* **100**, 443–448 (2003).
51. Zhou, H., Watts, J.D. & Aebersold, R. A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **19**, 375–378 (2001).
52. Oda, Y., Nagasu, T. & Chait, B.T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19**, 379–382 (2001).
53. Steen, H. & Mann, M. A new derivatization strategy for the analysis of phosphopeptides by precursor ion scanning in positive ion mode. *J. Am. Soc. Mass Spectrom.* **13**, 996–1003 (2002).
54. Medzihradszky, K.F. *et al.* The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* **72**, 552–558 (2000).
55. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
56. Oda, Y., Huang, K., Cross, F.R., Cowburn, D. & Chait, B.T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl Acad. Sci. USA* **96**, 6591–6596 (1999).
57. Stemmann, O., Zou, H., Gerber, S.A., Gygi, S.P. & Kirschner, M.W. Dual inhibition of sister chromatid separation at metaphase. *Cell* **107**, 715–726 (2001).
58. Ong, S.E., Kratchmarova, I. & Mann, M. Properties of ^{13}C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.*, in press (2003).
59. Blagoev, B. *et al.* A proteomics strategy to elucidate functional protein–protein interactions applied to EGF signaling. *Nat. Biotechnol.* **21**, 315–318 (2003).