ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Chromatography A, 1000 (2003) 589–608

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Review

Recent liquid chromatographic–(tandem) mass spectrometric applications in proteomics

Edwin P. Romijn, Jeroen Krijgsveld, Albert J.R. Heck*

Department of Biomolecular Mass Spectrometry, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands

Abstract

Conventional proteomics makes use of two-dimensional gel electrophoresis followed by mass spectrometric analysis of tryptic fragments derived from in-gel digestion of proteins. Although being a very strong technique capable of separating and visualizing hundreds of proteins, 2D-gel electrophoresis has some well-documented disadvantages as well. More recently, liquid chromatographic-(tandem) mass spectrometric techniques have been developed to overcome some of the shortcomings of 2D-gel electrophoresis. In this review we have described several recent applications of liquid chromatography-(tandem) mass spectrometry in the field of proteomics and especially in the field of membrane proteomics, quantitative proteomics and in the analysis of post-translational modifications.
© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Reviews; Proteomics; Liquid chromatography–mass spectrometry; Proteins; Peptides

Contents

1. Introduction ... 590
2. Conventional strategies in proteomics ... 591
   2.1. 2D gel electrophoresis ... 591
   2.2. Protein identification by mass spectrometry ... 592
      2.2.1. Peptide mapping, protein fingerprints ... 592
      2.2.2. Peptide sequencing by tandem mass spectrometry ... 592
   2.3. Drawbacks of the conventional approach in proteomics ... 592
3. LC-based approaches in proteomics ... 593
   3.1. Multi-dimensional LC–MS–MS ... 594
   3.2. High resolution LC–MS using accurate mass tags ... 595
4. Quantitative proteomics ... 596
   4.1. Absolute quantitation ... 596
   4.2. Strategies for isotopic labeling of proteins and peptides ... 597
   4.3. Labeling of peptides ... 597
   4.4. Cleavage labeling in ^{18}O water ... 598

*Corresponding author. Tel.: +31-30-253-6797; fax: +31-30-251-8219.
E-mail address: a.j.r.heck@chem.uu.nl (A.J.R. Heck).

0021-9673/03/$ – see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0021-9673(03)00178-X

4.5. ICAT labeling 598
4.6. Metabolic labeling 599
4.7. Isotope effect in LC 599
5. Post-translational modifications 600
5.1. Phosphorylation 601
5.1.1. Identification and enrichment of phosphorylated peptides 601
5.2. Glycosylation 603
5.2.1. O-GlcNAc modification 603
5.3. Ubiquitination 604
5.4. Other modifications 605
6. Summary 605
References 605

1. Introduction

The unraveling of the human genome [1,2] has been the starting point for addressing biological processes in a complete new perspective. With the limited set of genes available in humans, the complexity of biological processes is embedded in the complexity of the gene products (the proteins). The technological tools are now becoming available to study properties of proteins at a more comprehensive scale, including not only their expression patterns, but also the protein networks they form [3,4], the way they fold [5] and the way they are post-translationally modified [6–9]. Cell-, organism-, or for instance body fluid-wide analysis of proteins is nowadays termed proteomics. Such a comprehensive analysis of proteins is essential if we really want to make advances in both biomedical and biotechnology areas, as protein properties such as folding, post-translational modifications and protein networks are in many cases directly linked to diseases, such as Alzheimer's, BSE and rheumatoid arthritis [10–12].

One of the core components of proteomics is the ability to systematically quantify and identify every protein expressed in a biological cell or tissue. The technology for such proteome analysis involves separation science for the separation of proteins and peptides, analytical science for the identification and quantification of these biomolecules, and bio-informatics for data-management, including linking the proteomics data to data obtained via other genome-wide approaches. Proteomics is a multidisciplinary research activity wherein separation science and mass spectrometry play pivotal roles.

Proteomics is far more complex than genomics as it encompasses the characterization and functional analysis of all proteins that are expressed by the genome at a certain moment, under certain conditions. Since expression levels of proteins strongly depend on complex, but well-balanced regulatory systems the proteome, unlike the genome, is highly dynamic. This variation depends on the biological function of a cell, but also on signals from its environment. In biomedical research it becomes increasingly apparent that cellular processes, in particular in case of diseases, are determined by multiple proteins, and thus that it is important not to focus on one single gene product (one protein), but to study the complete set of gene products (the proteome). In this way the multi-factoral relations underlying certain diseases may be unraveled, opening new ways to drug-therapy.

In this review we focus on some of the newest analytical strategies in proteomic research. We briefly describe the conventional bio-analytical strategies used in proteomics, primarily based on 2D gel electrophoresis and mass spectrometry. Although these methods are extremely powerful and amenable for full automation, they also have some inherent drawbacks. Therefore, more and more alternative strategies are being developed to avoid 2D gels. We describe some of the most recent and relevant approaches, including several combinations of liquid chromatography with either single dimensional MS or tandem mass spectrometry. At the end of the review we particularly focus on two highly relevant areas in proteomics research, i.e. quantitative proteomics and post-translational modifications, and describe recent contributions both of LC–MS and LC–MS–MS to those fields.

2. Conventional strategies in proteomics

Isolation, separation and analysis of proteins is much more difficult than similar maneuvers in the DNA and RNA world. This is largely due to the much more heterogeneous nature of the proteins compared to nucleic acids. Additionally, the protein world lacks a PCR equivalent, which could be used for amplification. The conventional strategy for proteome analysis is outlined in Fig. 1. In general, proteomics starts with the isolation of the proteins from the biological matrix, such as fluids, cells or tissue. The procedures for these steps are highly dependent on the particular sample and/or organism of origin. It is important to separate the proteins from non-protein material, such as lipids, DNA, and RNA. For the separation and analysis of the proteins, conventionally two analytical methods have emerged, most often used in combination, namely two dimensional electrophoresis and mass spectrometry.

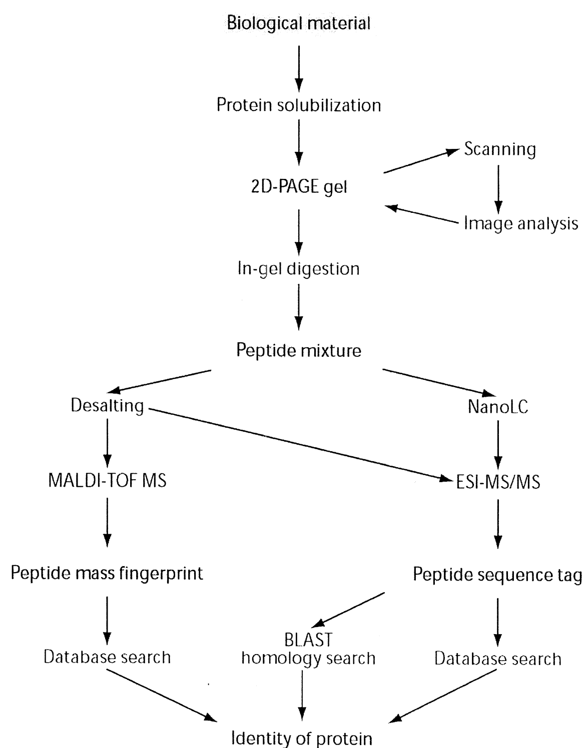


Fig. 1. Schematic overview of the conventional approach in proteome analysis based on 2D gel electrophoresis, mass spectrometry and data-base searches.

2.1. 2D gel electrophoresis

One of the main challenges in proteomics is the separation of the vast amount of different proteins originating from biological fluids or tissue. With genomes encoding often more than 30 000 genes, it may be expected that more than 300 000 proteins may be present even in simple organisms. In order to be able to study the hundreds of thousands of different proteins present in cells, these proteins must first be extracted and separated. To achieve this, two-dimensional (2D)-electrophoresis is at present still the most appropriate technique [13–17]. This technique separates proteins on the basis of their iso-electric point (iso-electric focusing, IEF) and molecular mass (gel electrophoresis, SDS–PAGE). Generally, in the first dimension the proteins are brought on a strip that contains an immobilized pH gradient. By applying an electric field over this strip the proteins will migrate over the strip until they reach the pH area on the strip where they will be neutral. Each protein therefore will be separated and focused on the strip at the position of its iso-electric point. In the second dimension proteins are separated on their size/mass. On the resulting two-dimensional gel each protein is present at a position that reveals it approximate p*I* and mass. To analyze and image these gels the proteins have to be visualized, which usually is achieved by staining them. Many different staining procedures can be used, including Coomassie blue staining, silver staining, and more recently staining by fluorescent dyes [18]. For expression proteomics a disadvantage of most staining techniques is that the magnitude of staining often depends on the nature of the protein, and provides usually only a limited dynamic range. Still imaging of the stained two-dimensional gels provides a very nice way to compare two proteomes, for instance for a stressed and normal state of a cell culture. Differential staining by using different fluorophores (i.e. fluorescent dyes) is a recent advancement, which increases the dynamic range of image profiling of protein expression patterns [18,19]. In summary, 2D gel electrophoresis is still one of the best techniques for separating and visualizing a large number of

proteins. Up to 10 000 different spots may be resolved on the best gels, especially when magnified zoom-gels are used [14].

2.2. Protein identification by mass spectrometry

Nowadays, an ordinary mass spectrometer can precisely determine the masses even of large proteins (~1 Da precision at 50 kDa). Since the sole, though precise mass of a protein gives no direct clue about its identity, in practice two distinct mass spectrometric techniques are employed for protein identification [20,21]. In the first method a peptide fingerprint of a protein is recorded, usually by MALDI-TOF mass spectrometry. In the second, slightly more complicated method, short amino acid sequences, so-called sequence tags, are determined by tandem mass spectrometry.

2.2.1. Peptide mapping, protein fingerprints

In the first of the two approaches, the protein spot to be identified is cut out of the gel and digested (in-gel) with a protease, most often trypsin, which specifically cleaves at arginine or lysine residues. The resulting peptide mixture is eluted from the gel, and analyzed by matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry, or alternatively by using electrospray ionisation mass spectrometry. Collectively, these peptide masses form a fingerprint, which is indicative for the protein concerned. This fingerprint is then compared to theoretically expected tryptic peptide masses for each protein entry in the database. The hits are generally ranked according to the number of peptides that match. Unique identification does not require the whole protein to be covered by the tryptic peptides. Usually 10–20% coverage is already sufficient. After a hit a second search is usually done to correlate the remaining peptides in the mass spectrum with the data obtained from the protein hit, taking into account possible post-translational modifications. The great precision of mass determination, which may be achieved by modern mass spectrometers, allows the discrimination of even highly homologous proteins, or proteins which only differ by a single amino acid. Generally, peptide finger printing is still the most rapid and efficient method for protein identification. As identification occurs via consultation of protein and genome databases, it may be apparent their increasing comprehensiveness greatly aids in protein identification.

2.2.2. Peptide sequencing by tandem mass spectrometry

If the peptide mass fingerprint does not allow an identification of the protein (for example because insufficient peptides have been generated and/or the fingerprint is not unique), then it is possible to determine a small piece of amino-acid sequence from one or more of the peptides. This occurs by fragmentation of the peptide in the mass spectrometer and mass-analysis of resulting fragments. Although these fragmentation patterns maybe quite complicated they generally allow the determination of partial sequences. With this partial sequence, possibly in combination with the peptide-fingerprint already obtained, the chance of a unique hit in the database is considerably enhanced. With one or two of these short sequence tags (often no more than five amino acids), it is often possible to unambiguously identify a protein. Should a match not be found in a database, additional peptides have to be mass spectrometrically sequenced in this manner. Determination of the total sequence of a protein still requires considerably more effort.

2.3. Drawbacks of the conventional approach in proteomics

Although, the general conventional procedures for proteomic analysis, described above, allow the analysis of a large number of proteins there are many inherent drawbacks. These drawbacks are related to (a) the limited number of proteins that can be visualized on 2D gels, (b) the limited quantitative validity and dynamic range of 2D gel image analysis, (c) the incapability of 2D gels to handle more basic and very large proteins (e.g. membrane proteins) and (d) the reproducibility of the 2D gels. Additionally, the whole conventional procedure is quite elaborate.

It has been shown for instance that the number of 2-D PAGE ‘‘spots’’ identified using the conventional 2D gel–MALDI-TOF MS approach are predominantly the most highly abundant ones, based on codon bias analysis [22]. The expected number of proteins in a proteome is much higher than the

protein spots observed on a 2D gel. The expected number of proteins can even be much higher than the number of genes as a single gene can give rise to multiple proteins due to co- and post-translational modifications, degradation intermediates and alternative splicing products. Many gene products are expressed at such low levels that their detection is precluded by the conventional 2D gel based approach, unless the proteome is selectively enriched. Much effort is being put into procedures to enrich the proteome for different classes of proteins [23–28] or selected protein networks [3]. With optimized enrichment procedures sub-proteomes can be selected, reducing the number of proteins to 10–100, which makes them more amenable for gel-based analysis.

3. LC-based approaches in proteomics

The above described conventional proteomic strategies, accomplished by the combination of two-dimensional gel electrophoresis (2DE) and MALDI-TOF mass spectrometry (MS), are nowadays more and more replaced by methods that do involve protein or peptide pre-separation by using high efficiency capillary separation techniques, such as liquid chromatography and/or electrophoresis. Capillary liquid chromatography separation efficiencies have been dramatically increased in recent years, by improving amongst others the packing material, the column manufacture and the operating pressures. Efficiencies greater than 10^5 plates/column can now be achieved with capillary columns packed with 1.5 μm particles, within a total separation time of half an hour [29–31]. When capillary liquid chromatography is combined with tandem mass spectrometry (LC–MS–MS) many of the limitations of 2D gel electrophoresis for proteome analysis can be overcome. The implementation of, in particular, nanoflow liquid chromatography mass spectrometry offers unique opportunities for speed, sensitivity and automation of proteomics research [32]. It is expected that nano-LC, combined with tandem mass spectrometry and database searching, will soon dominate the field of protein identification. The analysis of protein digests is typically performed using LC columns with 50–100-μm diameters, requiring the delivery of solvent gradients at nanolitre per minute flow-rates. This has been typically achieved using generic HPLC pumping systems for the delivery of microlitre per minute gradients that are either flow-split or sampled. At the end of the capillary often a tip is pulled with an inner diameter of ~5 μm. Typically, the total peptide mixture is first loaded onto a nanocolumn (usually 75 μm internal diameter) containing reverse phase C_{18} material and then eluted by using a gradient directly into a tandem mass spectrometer. The peptides are eluted at a slow flow-rate, typically 100–200 nl per minute, and the elution time of each peptide is ~10–30 s. The peptides elute off the column into the ionization source of the mass spectrometer. First a mass spectrum in the survey scan mode of the intact peptides can be obtained. In the data-dependent acquisition mode, the instrument can be set to automatically fragment and collect MS–MS data on any number of peptides observed in the MS spectrum based on their intensity, m/z value or charge state. Using such an LC–MS–MS approach, separations have been achieved with peak capacities around 10^3 in less than 3 h for even complex cellular lysate enzymatic digests [33,34].

In principle, using LC–MS–MS in proteomics, quite a few different strategies may be taken. In one approach, close to the conventional one, 2D gel electrophoresis is still employed for the separation of the intact proteins. Subsequently, the cut-out spots are digested and analyzed by LC–MS or LC–MS–MS. This approach may be complementary to a conventional peptide mapping/MALDI TOF approach as usually different sets of peptides may be more abundant/present in mass spectra obtained by MALDI and ESI. Such a parallel MALDI peptide mapping and ESI LC–MS approach therefore typically enhances the protein coverage. Such an enhancement of protein coverage is often not required for protein identification, but may be relevant in the analysis of protein modifications and/or mutations.

Several 2D gel independent LC–MS–MS approaches have been introduced to overcome some of the inherent disadvantages of 2D gels. In one approach the proteins in the total proteome are only separated and resolved by molecular mass using single-dimensional (1D) gels. Subsequently, this 1D gel is cut into pieces, all proteins in such a band are

digested and the mixture of peptides are analyzed by LC–MS and/or LC–MS–MS [35]. This approach provides an intermediate form between analyzing the very complex large peptide mixture obtained when digesting all proteins of a lysate and the single protein digested when using a 2D gel. As advantages over the 2D gels, a 1D gel-based approach is less elaborate. Additionally, very large and basic proteins are easier to handle using just single-dimensional gels.

3.1. Multi-dimensional LC–MS–MS

In a third approach the whole cell lysate is digested chemically or by a protease. This generates a very complex set of peptides, beyond the separation capacity of 1-D separation techniques. For the analysis of such complex mixtures, several multi-dimensional separation techniques have been introduced. These include several combinations of chromatography, such as size exclusion, reversed-phase and ion-exchange, but also combinations of chromatography and capillary electrophoresis. All these multi-dimensional chromatography techniques have much higher separation capacities than 1-D methods and can potentially be automated for high-throughput experiments. The combination of the high separation capacity of multidimensional LC and the powerful peptide characterization ability of tandem mass spectrometry allow the analyses of very complex protein samples.

Especially, the pioneering work by Yates' group has revealed the power of multi-dimensional chromatography coupled to MS–MS for proteomics. They introduced the rapid and large-scale proteome analysis of total cell lysates by multidimensional protein identification technology, termed by them MudPIT [36,37] (see Fig. 2). These technologies employ for instance a biphasic column with a section of reversed-phase material flanked by strong cation-exchange resin. They applied this MudPIT technology to the proteome analysis of *Saccharomyces cerevisiae*, which yielded one of the largest proteome analyses to date, as far as the number of detected/identified proteins is concerned [38]. A total of 1484 proteins were detected and identified. Categorization of these hits demonstrated the ability of this technology to detect and identify proteins rarely seen in

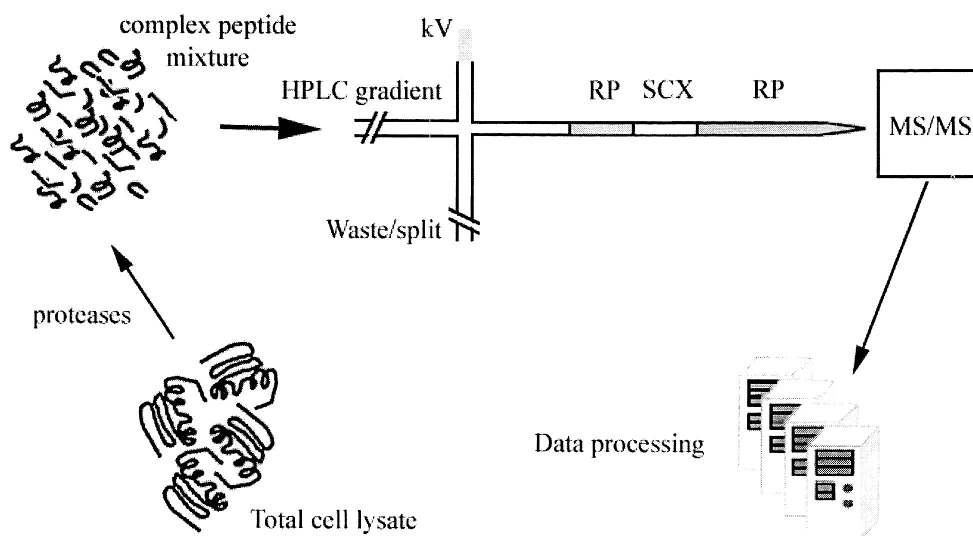


Fig. 2. Schematic overview of the three-phase multidimensional protein identification technology (MudPIT). A total cell lysate is digested by a protease (i.e. trypsin). The complex peptide mixture is loaded on a capillary column. The HPLC gradient is delivered to a junction and part of the flow is split off through the column to get a flow of a couple of hundreds of nl/min. The peptides are separated by a three-phasic column, with in line a reversed-phase (RP)–strong cation-exchange (SCX) and another reversed-phase, before they enter the ionization source of the mass spectrometer. Peptide elutes in a time window of several seconds allowing the mass measurements and the determination of peptide sequences by tandem mass spectrometry. These data are used for protein identification using database searches.

proteome analysis, including low-abundance proteins like transcription factors and protein kinases. Even more recently McDonald et al. [39] reported the use of a third dimension/phase adding an additional section of RP material behind the first RP and SCX (see Fig. 2). Although, this extra RP dimension of separation indeed allowed the identification of even more proteins the whole data collection time had to be significantly extended.

Compared to the conventional 2D gel-based approach multi-dimensional LC has also proven to be profitable for the analysis of membrane proteins. Membrane proteins are usually underrepresented on conventional 2D gels. This can be attributed to several factors, although their intrinsic low solubility is the most important one. Keeping to 2D gel electrophoresis for protein separation, several strategies have been introduced to improve the proteomic analysis of membrane proteins. These include prefractionation [40], solubilization of membrane proteins using nonionic/zwitterionic detergents [40–43] and/or organic solvent extraction of peptides from gels [44] and specialized immobilized pH gradient gels in the range $10<pI<12$ [14,45]. Although, all these described methods improve the analysis of membrane/alkaline proteins significantly, there seems yet not to be a simple ideal approach to membrane proteins in 2D gel-based proteomic analysis. Therefore, in particular in the analysis of membrane proteins there might be a need for alternative approaches for protein separation. Using MudPIT strategies the group of Yates identified 131 proteins in yeast with three or more predicted transmembrane domains [34,38]. MudPIT is useful for proteome analysis in general and may be specifically applied to integral membrane proteins to obtain detailed biochemical information on this ‘‘unmanageable’’ class of proteins.

3.2. High resolution LC–MS using accurate mass tags

Although the measurement of sequence tags by LC–MS–MS techniques is a very powerful tool in proteomics for protein identification, alternatives have been introduced which omit the tandem mass spectrometry step and make use of the fact that many tryptic peptides have unique masses. Such methods have been pioneered by the group of Smith [46–51] and make use of capillary LC or capillary electrophoresis combined with high resolution Fourier transform ion cyclotron resonance mass spectrometry.

When we consider only peptides with a mass between 500 and 4000 Da there are theoretically more than 10^{50} peptide sequences possible. However, when a whole theoretical proteome is in silico digested, using the specific trypsin cleavage sites, a much smaller limited number of peptides is generated. For instance, for yeast with its ~6200 genes, there are theoretically only 200 000 potential tryptic peptides. Some of these 200 000 peptides have identical masses, although also many peptides exist that have a unique mass. When one could measure this mass extremely accurately the mass measurement of such a single unique mass peptide would be sufficient to identify a particular protein. Fourier transform ion cyclotron mass spectrometry is at present still the method of choice for the measurements of accurate masses, and allows the mass measurements of a peptide with a mass below 2500 Da within 0.5 ppm (or 1 mDa). If we consider this mass accuracy achievable, approximately one third of these 200 000 peptides have a unique mass, and if we could identify them all, one would cover 98% of all predicted open reading frames (in yeast). This is most probably the most reductionist view of mass spectrometry-based proteomics. Primarily, Smith and co-workers [46–48] have pioneered this strategy, although others have used similar approaches [52]. This approach has as advantages that it extends the sensitivity, dynamic range, comprehensiveness, and throughput of proteomic measurements. In recent studies on the proteome of the ionizing radiation resistant bacterium *Deinococcus radiodurans* [49–51] they characterized peptide mixtures of significantly more than 10^5 components with mass accuracies of <1 ppm. Fig. 3 shows a two-dimensional display of all peptides analysed by capillary liquid chromatography coupled to an FT-ICR mass spectrometer, in which more than 50 000 putative peptides were detected from a tryptic digest of proteins harvested from *Deinococcus radiodurans*. The strength of the LC–FT-ICR combination allowed also an overall dynamic range of 10^4–10^5. Using solely the AMTs, more than 60% of the potentially

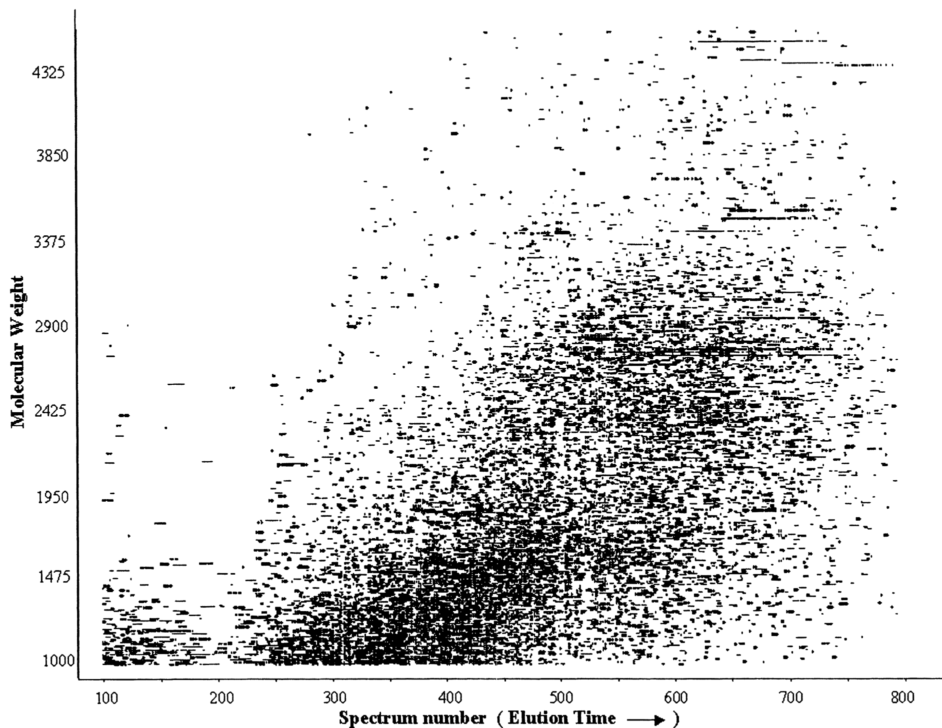


Fig. 3. Two-dimensional display of peptides analysed by capillary LC coupled to an FT-ICR mass spectrometer, in which more than 50 000 putative peptides were detected from a tryptic digest of proteins harvested from *Deinococcus radiodurans*. Adapted from Ref. [51] with permission from the publisher.

expressed proteins could be identified. They also demonstrated that this approach might be easily combined with stable isotope labeling (see below), making it amenable for quantitative expression proteomics. Using this approach they were able to investigate important biological processes, such as stress response and DNA repair quantitatively at the proteome level [51]. Although at the moment it is somewhat restricted to (expensive) high resolution mass spectrometric instrumentation, the use of accurate mass tags in combination with LC–FT-ICR–MS is a very powerful additional tool in proteomic research.

4. Quantitative proteomics

One of the major challenges in proteomics is to quantify relative expression levels of individual proteins. The traditional way to do this is by 2D electrophoresis and comparison of densities of corresponding spots in different gels as described above. This provides only a rough estimate, because it is critically dependent on sample processing, staining efficiency of individual gels and subsequent spot detection. A more recent approach makes use of dual labeling of protein samples prior to electrophoresis with fluorescent dyes emitting at different wavelengths [18,19,53]. Since samples can be mixed and run in the same gel, some of the earlier limitations can be overcome. However, samples still have to be processed in parallel resulting in possible introduction of errors.

4.1. Absolute quantitation

More recently, alternative approaches have been developed in which relative quantitation is determined by MS. These methods are based on the direct/linear relationship between the amount of peptide/protein in a sample and its peak height in a mass spectrum [54]. However, direct quantitation of

proteins and peptides by MS is difficult if not impossible because the absolute intensities of the detected ions are determined by their physico-chemical properties. Therefore internal standards would be needed with chemical properties very similar to the analyte. To this end, the technique of isotope dilution has proven highly suitable, which involves the spiking of the sample with the analyte modified with a stable isotope (^{2}H, ^{13}C, ^{15}N). The intensity ratio of the internal standard to the analyzed compound is then used for quantitation. If the absolute concentration of the standard is known, then the quantity of the analyte can be determined accurately. This is shown in a recent study [55] in which the phosphorylation status of a protein in a complex mixture was quantified. Such an approach would in principle be possible for all proteins in a sample, but would require the synthesis of all peptides present in labeled form.

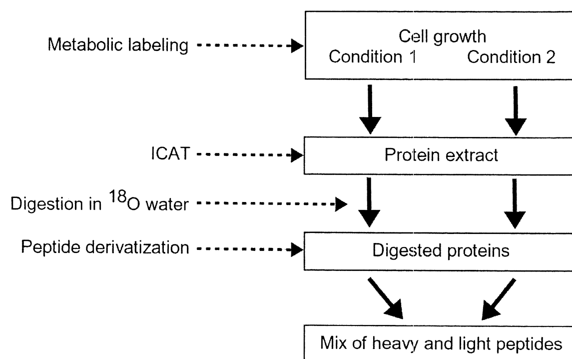


Fig. 4. Various strategies for isotopic labeling of proteins and peptides. Stable isotope labels can be introduced in proteins at various stages of sample preparation. In a typical experiment, labeled and unlabeled samples should be combined directly after the labeling step. Metabolic labeling is the earliest possible point of entry and allows combination of samples directly after cell harvesting, thereby minimizing errors due to sample handling. Later entry points are provided by ICAT, digestion in ^{18}O water, and labeling after proteolytic digestion. Adapted from Ref. [36].

4.2. Strategies for isotopic labeling of proteins and peptides

For proteome-wide studies, relative rather than absolute quantitation is attained by derivatizing all proteins/peptides in one sample, which are then mixed with an unlabeled sample. Every labeled peptide in the mixture then serves as an internal standard for the corresponding unlabeled peptide. Primary requirements of this approach in proteome-wide studies is that all proteins/peptides in a sample are labeled, and that all pairs of labeled and unlabeled analytes can be both quantified and identified. Several labeling strategies have been developed over the past few years. Most of these procedures follow the sample flow from cell via extracted protein to (total) protein digest, but differ in the point of entry of the isotope label (Fig. 4).

Basically, labeling strategies can be divided into four approaches: labeling of peptides by derivatization with isotope-containing reagents, generation of peptides and concomitant labeling in ^{18}O-containing water, protein derivatization with isotope-coded affinity tags (ICAT), and metabolic labeling (Fig. 4). Although the final stage in all these procedures is a mixture of labeled and unlabeled peptides, they differ in the moment the actual mixing occurs, which is directly after the labeling step. Therefore, derivatization downstream in the procedure necessitates parallel sample preparation, while early introduction of the isotope, such as metabolic labeling, allows combination of the samples even prior to protein extraction.

4.3. Labeling of peptides

The general aim of any labeling strategy in comparative proteomics will be the derivatization of all proteins/peptides in a sample. Entities present in virtually all peptides after tryptic digestion are a primary amine and a carboxyl group at the N- and C-termini, respectively, and both are available for derivatization. Differential isotope labeling has been performed by acylation of N-terminal amines using *N*-hydroxysuccinimide or its trideuterioacetylated analogue [56,57]. Nicotinyl-*N*-hydroxysuccinimide has been used for the selective modification of N-termini, without affecting ε-amines in lysine [58]. Furthermore, the acidity of the modifying group caused preferential formation of b-ions in tandem mass-spectrometry aiding in spectrum interpretation [58].

4.4. Cleavage labeling in ^{18}O water

C-terminal labeling can be performed by tryptic digestion in $H_2^{18}O$ [59–62]. The ^{18}O atom is incorporated in the C-terminal carboxylic group of the peptides, and hence results in a mass increase of 2 Da compared to the non-derivatized peptide. This approach can be very powerful because labeling is universal and high labeling efficiencies can be obtained. A potential complication is that trypsin can continue to exchange ^{18}O into the peptide at the second oxygen, resulting in mixtures of isoforms and necessitating deconvolution of spectra [59,63]. Other parameters recognized to affect labeling efficiency are back-exchange due to hydrolysis, effect of pH, the nature of the peptide, and relative concentration of $H_2^{18}O/H_2^{16}O$ [63].

4.5. ICAT labeling

A different approach is taken by Gygi et al. [64] who developed an isotope coded affinity tag (ICAT) (Fig. 5). Tags of different isotope composition (heavy and light) are reacted with two protein populations to quantitatively differentiate between proteins altered in abundance. The ICAT reagent itself consists of a thiol-specific group, a linker region containing either eight hydrogen atoms (d0, light) or eight deuterium atoms (d8, heavy), and a biotin group. A protein sample from one source is reacted with the d0 version of the tag, another with the d8 version (Fig. 5). These samples are then combined and digested with a proteolytic enzyme. The biotin moiety is then used to selectively capture reacted peptides using avidin-chromatography. Since the method specifically targets cysteine-containing peptides this results in a substantial reduction in sample complexity so that a reduced number of peptides needs to be analyzed. Nevertheless, subsequent fractionation by reversed-phase capillary LC is needed to be able to analyze a significant portion of the mixture by MS. A combined on-line approach (LC–MS–MS) can be used for identification by peptide fragmentation, and the abundance ratio of peptides derivatized by either label can be determined by integration of peaks that are separated by 8 Da. To obtain higher quality data for quantification

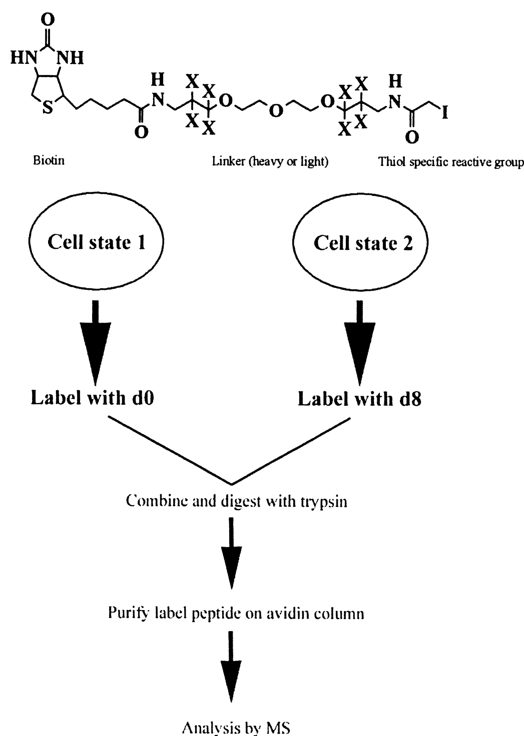


Fig. 5. Application of isotope coded affinity tags (ICAT). Two protein pools are reacted with the heavy (X = D) and light form of the label (X = H), respectively. Both pools are mixed, proteolyzed, and labeled peptides are isolated by avidin affinity chromatography. The resulting fraction is then analyzed by reversed-phase LC–MS–MS.

the mass spectrometer could also be operated in single MS mode (LC–MS).

Being developed only recently, there is a limited number of large-scale studies applying ICAT. One example shows the differential analysis of microsomal fractions of human cells which resulted in the identification and quantification of 491 proteins [65]. Further studies may be facilitated by implementation of software tools for mining the large amount of data that is generated [66]. At the same time, modifications of the original tag may allow more sensitive probing of low-abundance proteins, as shown in a recent report in which the ICAT reagent was linked to a solid-phase via a photocleavable linker [67].

Being intended for LC-based approaches, separation of ICAT-tagged proteins by two-dimensional gel electrophoresis was problematical indeed because

of smearing of spots [68]. This was shown to be due to incomplete or heterogeneous labeling, which could be alleviated after optimizing the labeling protocol [68,69]. Although identification and quantification after gel-based separation appears feasible, such an approach disregards the presence of the biotin moiety. Furthermore, it remains unclear what the influence of sub-optimal labeling efficiency is in the original LC-based analyses.

4.6. Metabolic labeling

Metabolic labeling offers the most comprehensive way of proteome coverage. Also, it is the earliest timepoint to introduce labels without the need of any in vitro derivatization steps (Fig. 4). The advantage is that cell/organisms to be analyzed can be combined even before protein extraction, thereby eliminating any variation due to subsequent sample handling. The biggest limitation of this method is that it can only be applied to organisms able to grow in defined media to which the desired label can be added. This is possible for yeast [70] and bacteria [71,72], and has also been applied to cell cultures [71,73]. Recently this has been extended to multicellular organisms by the complete labeling of the nematode *C. elegans* [74]. In all these organisms there is basically no restriction in the type of isotope used: labeling has been achieved by complete labeling by ^{15}N [70,71], but could be done using ^{13}C equally well. A cheaper approach is the incorporation of one or two labeled amino acids (e.g. deuterated leucine, methionine) [73]. This offers the advantage that the fixed increase of mass is a direct measure of the number of incorporated labeled residues, which could even help to confirm identification [75].

Compared to the ICAT approach, no reduction in sample complexity is obtained. On the other hand, higher protein coverages can be achieved with no bias against cysteine-free proteins.

Metabolically labeled proteins can be analyzed by 2-D electrophoresis since labeled and unlabeled proteins co-migrate. However, when it is preferred to avoid gels, LC–MS–MS is a highly suitable method for analyzing such mixtures after digestion. Co-eluting peptides would ideally be quantified and identified in the same run; otherwise consecutive runs of the same sample can be performed. In the first run peptides can be quantified, the second run then is used for subsequent identification of selected peptides meeting preset criteria of up- or down-regulation [36]. A major task for the near future will be how to distinguish up- or down-regulated peptides among many thousands of non-varying peptides, followed by their identification. Preferably this should occur in one single experiment.

4.7. Isotope effect in LC

The advocated advantage for the use of stable isotopes is their identical physico-chemical behaviour compared to non-labeled peptides. However, isotope effects may become apparent during chromatographic separations, resulting in the successive elution of differently coded peptides. Isotope effects may complicate the interpretation of the data and may even lead to the inability to recognize peak pairs. This would especially be true for highly complex samples with many peptides eluting simultaneously. More seriously, data can even become unreliable when peptides are differentially ionized because of changing matrix conditions.

Isotope effects have been explored in considerable detail by Zhang et al. [76–78]. They as well as others have shown that the effect is mainly apparent for deuterated peptides, and has been recognized for ICAT [64,76] but also various other types of labels [77]. The effect was not observed for ^{13}C-type labels [78].

The impact of the effect depends on how data are collected. Ideally, this would be an automated process in which first the abundance ratio of coeluting peptides is determined, followed by fragmentation if preset criteria are met. When peptides start to resolve (even partially), no reliable ratio could be calculated since the outcome would vary across the peak (Fig. 6) [76]. In such cases quantitation can still be done by integration of total isotope envelopes. In an effort to determine which factors are involved in the isotope effect of deuterium-based labels, Zhang et al. have found a direct relationship between this effect and the size of the labeling agent (relative to the size of the peptide). Furthermore, the isotope effect was enhanced with increased deuterium content, especially if they were positioned near hydrophobic domains

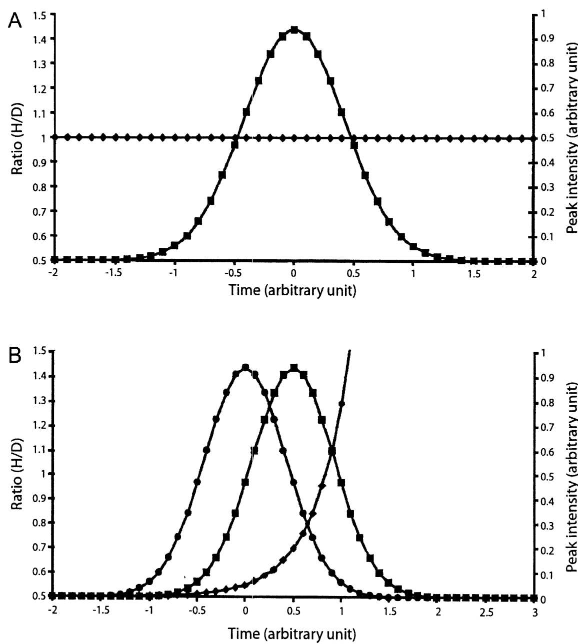


Fig. 6. Simulation of isotope effects during chromatographic separation. Panel A, the abundance ratio (♦) of co-eluting peptides will be the same at every timepoint. Panel B, when peaks start to resolve, the abundance ratio will vary continuously across the peak. Adapted from Ref. [76] with permission from the publisher.

in the peptide, which are the most likely regions to interact with the reversed-phase resin [77].

5. Post-translational modifications

A thorough understanding of cell architecture, regulation and dynamics, makes it necessary to not only identify the gene products but also to investigate the spatial and temporal protein distributions, composition and dynamics of protein complexes and post-translational modifications of proteins [79,80]. A unique feature of proteome analysis is that post-translational modifications can be investigated at a more global level [80,81]. It is known that the expressed products of a single gene represent a protein population that can contain large amounts of micro-heterogeneity (Fig. 7). Each state (e.g. another phosphorylation, glycosylation, ubiquitination, acetylation, lipidation, etc.) of a protein adds a large amount of diversity to the expression profile of that protein. More than 100 modification types are described and very likely more will be found in the future [82]. As an example we just like to mention that it has been observed that the prion protein

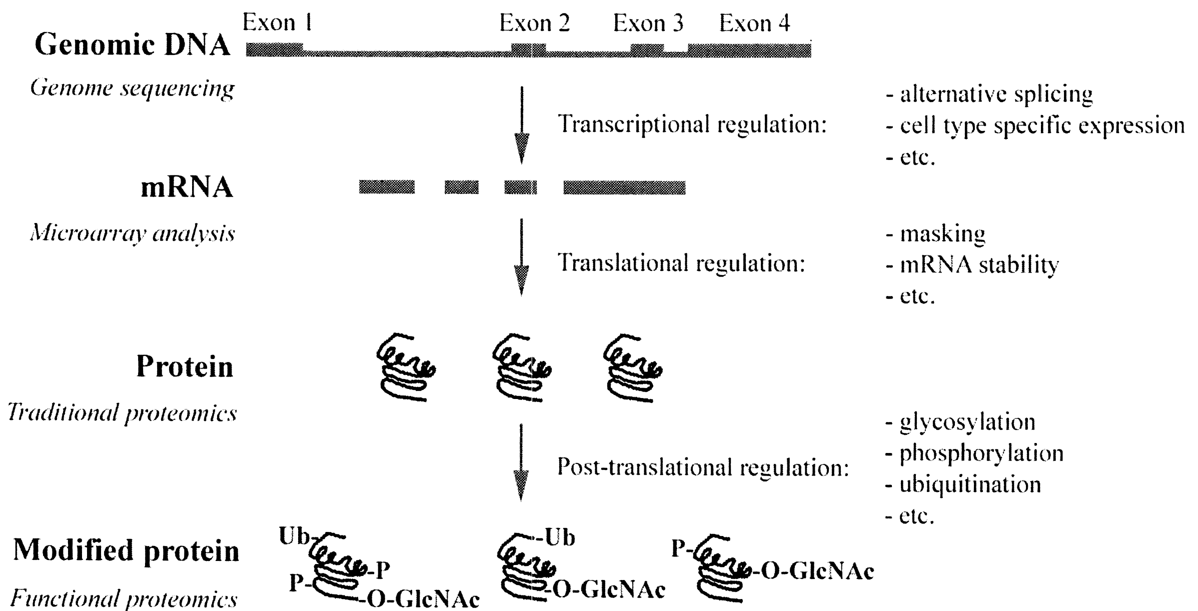


Fig. 7. A single gene can give rise to a lot of different gene products. Complexity increases dramatically as you go from genomic DNA to modified proteins.

extract from the human brain may contain already more than 60 isoforms, mainly due to heterogeneous glycosylation [83]. Clearly, post-translational modification of proteins is an event that has a drastic effect on the complexity of the proteome. We focus here on a few important post-translational modifications that have been studied in global proteomics studies using LC–MS–MS-based technologies.

5.1. Phosphorylation

Maybe the best-known and most important post-translational modification is protein phosphorylation. The study of protein phosphorylation has grown exponentially in recent years as researchers from various disciplines have come to realize that key cellular functions are regulated by the reversible phosphorylation and dephosphorylation of proteins on serine, threonine and tyrosine residues [84–87]. Approximately 2000 genes encoding protein kinases and probably half as many genes encoding protein phosphatases are estimated to exist in the human genome [88]. At present, only a small subset of eukaryote protein kinases and protein phosphatases has been characterized with respect to biological function and protein substrate specificity. To understand more about protein phosphorylation and dephosphorylation, it is necessary to identify the specific amino acid residues that become phosphorylated, because identification of these sites in proteins may reveal which protein kinase regulates the protein and thereby help elucidate the biological function and significance of novel phosphoproteins. A conventional method used to study protein phosphorylation employs radiolabeling with ^{32}P inorganic phosphate ($^{32}P_i$). To measure differences in relative abundances of phosphorylation, ^{32}P-labeled proteomes are resolved by two dimensional polyacrylamide gel electrophoresis (2-D PAGE) and the relative spot intensities are compared [89]. The use of $^{32}P_i$ to label proteins does not lend itself to high-throughput proteome-wide analysis due to safety issues with handling radioactive compounds and the associated contamination of analytical instrumentation.

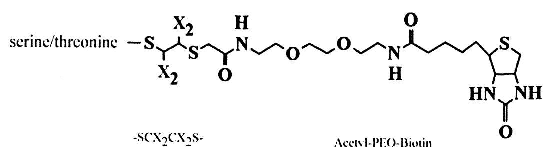
Several new methods depend on mass spectrometry for the analysis of phosphoproteins and have been described extensively [8,66,90–96]. All these methods rely on the fact that phosphorylated peptides have an increased mass of 80 Da for every phosphorylated serine, threonine or tyrosine. Phosphorylated peptides also have a different retention in reversed-phase LC when compared to their unphosphorylated counterparts and therefore LC may be used to separate them. In vivo protein phosphorylation is a very complex process, and therefore difficult to analyze. Phosphorylation is a dynamic process and usually only a fraction of the protein is phosphorylated (at a certain time point). Moreover, the negatively charged phosphogroup suppresses ionization (in positive ion mode) dramatically. Therefore, to be able to analyze protein phosphorylation enrichment strategies have been introduced, that involve modifying phosphoproteins with affinity tags. These chemical modifications are often performed in combination with stable isotope incorporation that allows for their subsequent identification and quantitative analysis by MS [70,97–99]. Below we describe these strategies and their applications in more detail.

5.1.1. Identification and enrichment of phosphorylated peptides

One of the primary difficulties in identifying phosphopeptides in complex mixtures is the fact that phospho-specific antibodies and metal affinity columns, that are widely used, typically result in co-isolation of nonphosphorylated species along with the phosphopeptides of interest. Fortunately, several new methods have been developed recently that provide tools for the specific enrichment of phosphopeptides from complex mixtures [97–101].

A first method was developed concurrently, and independently, by two groups and are essentially identical [97,98,100]. This approach to phosphoprotein enrichment and mapping is based on site-specific modification of phosphoseryl and/or phosphothreonyl residues. The phosphates are chemically replaced by biotinylated moieties. Base hydrolysis is used to induce β-elimination of phosphate from phosphoserine and phosphothreonine, followed by addition of 1,2-ethanedithiol (EDT) to the alkene. The resulting free sulfhydryls are coupled to biotin, allowing purification of phosphoproteins by avidin affinity chromatography. After elution of phosphoproteins and proteolysis with trypsin, enrichment

is carried out by a second round of avidin purification [98]. However, an undesired side effect involving side chains on cysteine and methionine residues can occur. To overcome this problem, the sample is first treated with performic acid, leading to oxidation of these residues, thereby inactivating them. Essentially the same strategy has been used with a labeled version of EDT (d0 and d4) for the quantitation of phosphorylation and is described later [100] (Fig. 8). It is basically the same as the ICAT strategy described earlier (Fig. 5), but specifically applicable to protein phosphorylation. The main disadvantage of this chemical modification method is that it is not applicable to tyrosine phosphorylation. Another problem is the low solubility of the thiol compound in water. However, because the reactions are performed in a single tube, this procedure is easier to



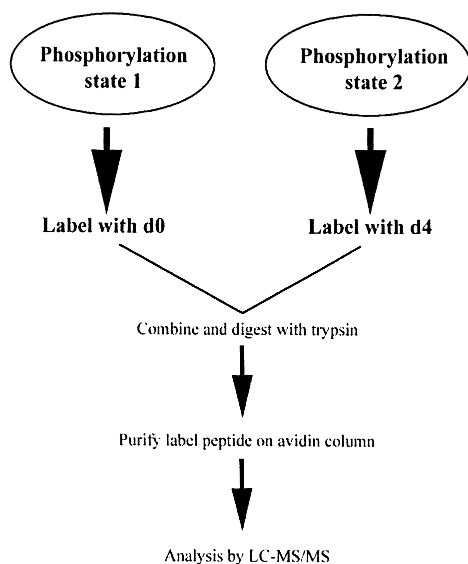


Fig. 8. Phosphoproteome analysis using phosphoprotein isotope-coded affinity tags. The first phase involves PhIAT labeling of the phosphoseryl and phosphothreonyl residues. In the second phase the proteins are digested and the sample is enriched for PhIAT-labeled peptides by avidin affinity chromatography. This enriched mixture is then analyzed by capillary reversed-phase LC–MS–MS.

perform than the method described below and losses resulting from multiple purification steps are minimized. A second strategy, developed by Zhou et al. [99] is applicable to phosphotyrosine-containing peptides in addition to those containing phosphoserine and phosphothreonine residues. The special feature of this method is ethyl carbodiimide (EDC) catalyzed addition of cystamine to phosphate moieties, which allows purification of phosphopeptides on glass beads containing immobilized iodoacetyl groups. The phosphopeptides are eluted by cleavage of phosphoroamidate bonds by TFA, a step that also regenerates the amino groups. Several chemical reactions and purification steps are needed using this approach, which possibly leads to substantial losses. As is generally the case with chemical modification-based approaches, both the above methods require substantial amounts of sample with the result that only abundant proteins are easily identified. Nevertheless, these approaches are promising and could be coupled to other fractionation steps to improve the overall recovery of low-abundance proteins.

An alternative method for phosphopeptide recovery makes use of miniaturized immobilized metal affinity chromatography (IMAC). The use of miniaturized IMAC columns for the enrichment of phosphopeptides exploits the high affinity of phosphate groups towards a metal-chelated stationary phase, especially Fe^{3+} and Ga^{3+}. IMAC has been successfully used in off-line and on-line formats for the detection of phosphopeptides using MS [93,102–105]. Because it is based on the presence of negatively charged phosphate groups, IMAC generally enriches for phosphorylated serine, threonine and tyrosine residues. A major disadvantage is that the specificity of this procedure is variable because of affinity for acidic groups (aspartic and glutamic acid) and to electron donors (e.g. histidine). In addition, multiply phosphorylated peptides are more enriched and the recovery of phosphopeptides appears to be largely dependent on the type of metal ion, column material and the elution procedure used. Recently, Ficarro et al. attained a much higher specificity employing esterification of acidic residues prior to IMAC enrichment [101]. In this way they sequenced hundreds of phosphopeptides from total yeast protein extracts. With further refinement, this technique may

offer the best hope for the identification and enrichment of phosphorylated peptides.

In all these phosphopeptide analysis strategies nanoLC–MS–MS is almost a prerequisite as it allows reducing the complexity of the peptide mixture significantly. Coupling of nanoLC systems to a mass spectrometer is valuable because separation of peptides by the upfront LC step decreases the ion suppression effect that usually obscures the observation/detection of phosphopeptides.

5.2. Glycosylation

Protein glycosylation is another post-translational modification that is essential for in vivo functions of particularly eukaryotic genes. Glycoproteins play a predominant role in cell–cell and cell–substratum recognition events in multicellular organisms. Hirabayashi and Kasai have asserted the importance to study glycans from a genome-wide viewpoint to understand the complexity of life [106]. Several reasons are given for this assertion: firstly, abundant and heterogeneous glycosylation occurs on all cells of all organisms. Glycan composition significantly reflects differences in cell types and states, e.g. species, individuals, tissues, developmental stages, etc.

Additionally, glycans have much higher potential to exert structural diversity than nucleic acids and proteins [107]. Having sufficient diversity is essential for biologically informative molecules. The number of saccharide components is relatively small, e.g. Glc, GlcNAc, Man, Gal, GalNAc, L-Fuc, Xyl, L-Ara and NeuAc, but the high variation in linkage and branching events, makes glycosylation probably the most complex post-translational modification.

Although glycosylation does not only occur in proteins, e.g. also glycolipids and proteoglycans are present in cells, we describe here pioneering research that has been performed on protein glycosylation using a proteomic approach. In eukaryotic proteins, there are three types of glycosylation, i.e. N-glycosylation, O-glycosylation, and GPI-anchoring [108]. To our knowledge there is not a single method that deals with all types of glycosylation. Only one method to study glycosylation will be reviewed here, which is a study that focused on O-GlcNAc modified glycoproteins [109].

5.2.1. O-GlcNAc modification

O-GlcNAc-modification has not been studied in great detail by mass spectrometry, although it has great biological relevance. O-GlcNAc modification is believed to be a regulatory modification. Most known O-GlcNAc-modified proteins are also phosphoproteins and these proteins are often members of reversible multimeric complexes. Methodology mainly used for identification of O-GlcNAc-modified proteins involve galactosyltransferase radiolabeling of O-GlcNAc moieties, generation of glycopeptides by proteolytic cleavage, purification by HPLC, and Edman protein sequencing. This procedure is tedious and suffers from the disadvantage of requiring relatively large amounts of pure, labeled glycopeptide. An elegant approach, which has been used to circumvent some of these limitations is proposed by Haynes and Aebersold [109] and enriches O-GlcNAc-modified peptides from complex mixtures using *Ricinus communis* (RCA I) affinity chromatography followed by galactosyltransferase labeling (Fig. 9). After enrichment using RCA I affinity chromatography, the O-GlcNAc modified peptides are separated by LC and analyzed by MS. This was achieved by the development of an MS–MS-based protocol that specifically detects peptides carrying the O-GlcNAc modification by the monitoring of a specific reporter ion and the identification of the protein from which the labeled peptide originated by CID and database searching. Both operations are performed automatically and in the same experiment. The procedure has femtomole-level sensitivity. The labeling of an O-GlcNAc group with galactose using galactosyl transferase produces a disaccharide substituent with a diagnostic mass of m/z 366. The removal of this disaccharide from a glycopeptide by CID requires much less energy than the fragmentation of the peptide backbone. It is possible to take advantage of this difference in required fragmentation energy by using a triple quadrupole mass spectrometer in a multistage experiment. This is achieved by employing precursor ion scanning at a relatively low collision energy level to detect those species which produce a diagnostic m/z 366 fragment, followed by CID at higher energy of the precursor ion to fragment the peptide backbone, and thus identify the underlying peptide. Apparently novel O-GlcNAc modified proteins were detected

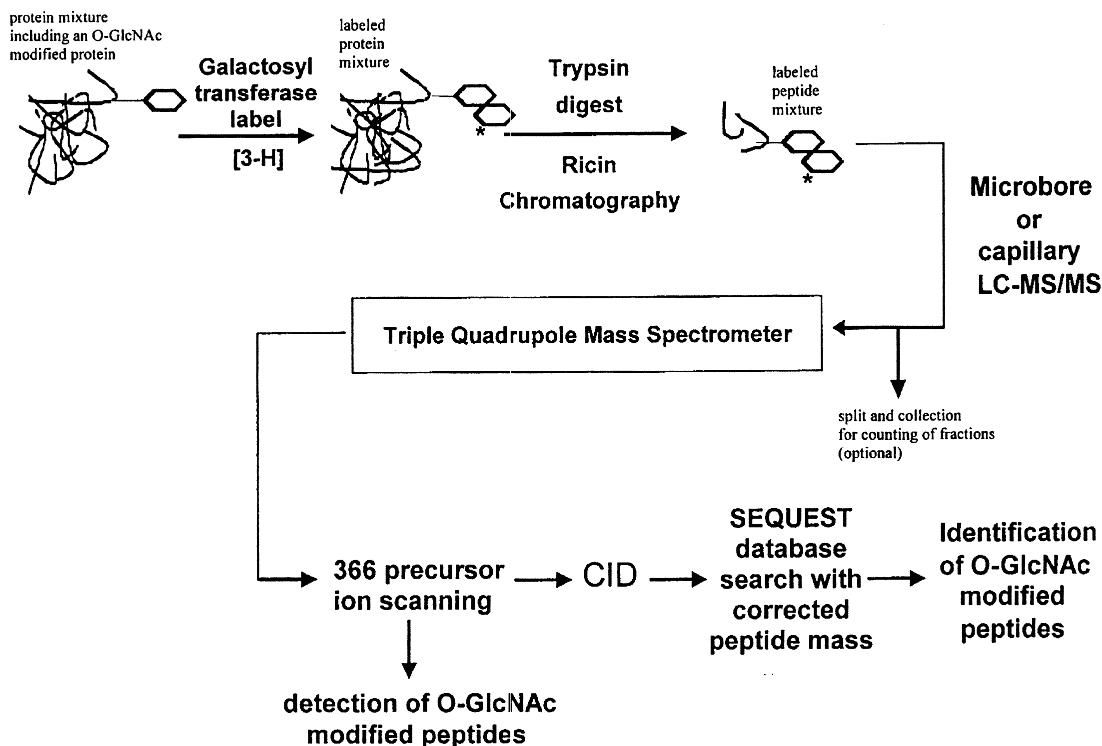


Fig. 9. Schematic diagram of the method described for the simultaneous detection and identification of O-GlcNAc-modified glycopeptides. The O-GlcNAc group is indicated by a hexagon and the labeled galactose group is indicated by a hexagon with an asterisk. Reprinted with permission from Ref. [109]. Copyright 2000 American Chemical Society.

directly from cellular lysates, but identification of the proteins was unsuccessful. The inability to identify detected proteins highlights two potential drawbacks of the method. Firstly, cellular lysates contain extremely complex mixtures of material, including protein, DNA, lipids, carbohydrates, and salts, which can interfere in both chromatographic separations and ionization in a mass spectrometer. Secondly, the database searching relies on matching of the spectra to a known sequence. In those latter cases, where a truly novel protein is encountered, database searches will not achieve a successful identification, irrespective of the quality of data involved. Although the method has its drawbacks, it has potential to produce meaningful results concerning the physical role of O-GlcNAc modified proteins.

5.3. Ubiquitination

Protein ubiquitination is among the most common of all post-translational modifications. Ubiquitin plays an essential role as a signal molecule for protein degradation in eukaryotes and as a regulator of protein activity and localization. While its biological importance is well acknowledged, not many reports appeared on the analysis of protein ubiquitination by LC–MS. Peng et al. have attempted to analyze protein ubiquitination by coupling multidimensional chromatography and tandem mass spectrometry (S.P. Gygi, personal communication). The strategy presented by Peng et al. is based on large-scale sequencing and characterization of protein ubiquitination in yeast by LC–MS–MS. Ubiquitin conjugates were purified from a yeast strain expressing His-tagged ubiquitin or a control strain by denaturing nickel affinity chromatography. Following trypsin digestion an ubiquitin conjugated protein contains a di-glycine residue of ubiquitin covalently attached to a lysine residue that is resistant to proteolysis. The resulting peptides were separated by

strong cation-exchange (SCX) chromatography. All fractions were sequentially analyzed by nanoLC–MS–MS.

A total of more than 1000 proteins were identified and from almost 10% of them the exact site of ubiquitination could be determined. Several proteins were found to be ubiquitinated and phosphorylated. Although this is a very elegant way to study post-translational modifications in a systematic manner, the method has disadvantages as well. A disadvantage of the strategy is that it heavily depends on the availability of His-tagged ubiquitin and that, from a total number of 1051, only 98 ubiquitination sites could be mapped. The exact ubiquitination sites of the other proteins remain unknown. Since the sequence coverage is almost never 100%, the method might even have missed ubiquitination sites.

5.4. Other modifications

The proteome world is full of post-translational modifications. Of the more than 100 modification types that have been described [82], only phosphorylation, glycosylation and ubiquitination have been illustrated above, since these modifications have been studied systematically using LC–MS–MS. Reports can be found in the literature whereby LC–MS or LC–MS–MS have been used to study other modifications, such as palmitoylation [110], lipidation [111] and sulfonation or modification by sialic acid [112]. Still, these reports and the ones we have addressed are just a starting point. As LC–MS–MS is at the moment probably one of the most appropriate tools for the comprehensive analysis of protein post-translational modifications we expect many more in the near future.

6. Summary

In this review we have focused on several recent applications of capillary separation technologies, and in particular liquid chromatography, coupled to (tandem) mass spectrometry in the field of proteomics. In the last decade capillary liquid chromatography separation efficiencies have been dramatically increased, and efficiencies greater than 10^{4-5} plates/column can now be achieved. With such separation power, which may even be increased by using multidimensional separation technologies, extremely complex samples such as total cell lysates, treated with a protease become directly amenable for analysis. This increases the speed and sensitivity of proteome analysis enormously. When capillary liquid chromatography is combined with tandem mass spectrometry (LC–MS–MS) many of the limitations of 2D gel electrophoresis for proteome analysis can be overcome. We have described the use of such LC–MS–MS technologies in the field of membrane proteomics, quantitative proteomics and in the analyses of protein post-translational modifications. It is expected that capillary (multidimensional) separation technologies combined with tandem mass spectrometry and database searching will soon dominate the field of protein identification in proteomics.

References

[1] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M.L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V.

Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N.N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J.F. Abril, R. Guigo, M.J. Campbell, K.V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, X. Zhu, Science 291 (2001) 1304.

[2] J.D. McPherson, M. Marra, L. Hillier, R.H. Waterston, A. Chinwalla, J. Wallis, M. Sekhon, K. Wylie, E.R. Mardis, R.K. Wilson, R. Fulton, T.A. Kucaba, C. Wagner-McPherson, W.B. Barbazuk, S.G. Gregory, S.J. Humphray, L. French, R.S. Evans, G. Bethel, A. Whittaker, J.L. Holden, O.T. McCann, A. Dunham, C. Soderlund, C.E. Scott, D.R. Bentley, G. Schuler, H.C. Chen, W. Jang, E.D. Green, J.R. Idol, V.V. Maduro, K.T. Montgomery, E. Lee, A. Miller et al., Nature 409 (2001) 934.

[3] A.C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, A.M. Michon, C.M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.A. Heurtier, R.R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, G. Superti-Furga, Nature 415 (2002) 141.

[4] Y. Ho, A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sorensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W. Hogue, D. Figeys, M. Tyers, Nature 415 (2002) 180.

[5] C.M. Dobson, Trends Biochem. Sci. 24 (1999) 329.

[6] W. Shou, R. Verma, R.S. Annan, M.J. Huddleston, S.L. Chen, S.A. Carr, R.J. Deshaies, Methods Enzymol. 351 (2002) 279.

[7] M. Mann, S.E. Ong, M. Gronborg, H. Steen, O.N. Jensen, A. Pandey, Trends Biotechnol. 20 (2002) 261.

[8] D.T. McLachlin, B.T. Chait, Curr. Opin. Chem. Biol. 5 (2001) 591.

[9] M.J. MacCoss, J.R. Yates 3rd, Curr. Opin. Clin. Nutr. Metab. Care 4 (2001) 369.

[10] G.V. Johnson, J.A. Hartigan, J. Alzheimers Dis. 1 (1999) 329.

[11] G.V. Johnson, S.M. Jenkins, J. Alzheimers Dis. 1 (1999) 307.

[12] Z. Zhou, H.A. Menard, Curr. Opin. Rheumatol. 14 (2002) 250.

[13] A. Gorg, C. Obermaier, G. Boguth, W. Weiss, Electrophoresis 20 (1999) 712.

[14] A. Gorg, C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, W. Weiss, Electrophoresis 21 (2000) 1037.

[15] T. Rabilloud, Proteomics 2 (2002) 3.

[16] T. Rabilloud, Anal. Chem. 72 (2000) 48A.

[17] T. Rabilloud, Methods Mol. Biol. 112 (1999) 297.

[18] W.F. Patton, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 771 (2002) 3.

[19] M. Unlu, M.E. Morgan, J.S. Minden, Electrophoresis 18 (1997) 2071.

[20] M. Mann, R.C. Hendrickson, A. Pandey, Annu. Rev. Biochem. 70 (2001) 437.

[21] J.R. Yates 3rd, J. Mass Spectrom. 33 (1998) 1.

[22] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, Proc. Natl. Acad. Sci. USA 97 (2000) 9390.

[23] A.S. Nouwens, S.J. Cordwell, M.R. Larsen, M.P. Molloy, M. Gillings, M.D. Willcox, B.J. Walsh, Electrophoresis 21 (2000) 3797.

[24] R.S. Taylor, C.C. Wu, L.G. Hays, J.K. Eng, J.R. Yates 3rd, K.E. Howell, Electrophoresis 21 (2000) 3441.

[25] M.F. Lopez, B.S. Kristal, E. Chernokalskaya, A. Lazarev, A.I. Shestopalov, A. Bogdanova, M. Robinson, Electrophoresis 21 (2000) 3427.

[26] A.W. Bell, M.A. Ward, W.P. Blackstock, H.N. Freeman, J.S. Choudhary, A.P. Lewis, D. Chotai, A. Fazel, J.N. Gushue, J. Paiement, S. Palcy, E. Chevet, M. Lafreniere-Roula, R. Solari, D.Y. Thomas, A. Rowley, J.J. Bergeron, J. Biol. Chem. 276 (2001) 5152.

[27] M.F. Lopez, Electrophoresis 21 (2000) 1082.

[28] M. Geng, X. Zhang, M. Bina, F. Regnier, J. Chromatogr. B Biomed. Sci. Appl. 752 (2001) 293.

[29] J.E. MacNair, K.C. Lewis, J.W. Jorgenson, Anal. Chem. 69 (1997) 983.

[30] J.E. MacNair, G.J. Opiteck, J.W. Jorgenson, M.A. Moseley 3rd, Rapid Commun. Mass. Spectrom. 11 (1997) 1279.

[31] J.E. MacNair, K.D. Patel, J.W. Jorgenson, Anal. Chem. 71 (1999) 700.

[32] H.D.H. Meiring, E. van der Hove, G.J. ten Jong, J. Sep. Sci. 25 (2002) 557.

[33] Y. Shen, R. Zhao, M.E. Belov, T.P. Conrads, G.A. Anderson, K. Tang, L. Pasa-Tolic, T.D. Veenstra, M.S. Lipton, H.R. Udseth, R.D. Smith, Anal. Chem. 73 (2001) 1766.

[34] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates 3rd, Nat. Biotechnol. 17 (1999) 676.

[35] J. Rappsilber, S. Siniossoglou, E.C. Hurt, M. Mann, Anal. Chem. 72 (2000) 267.

[36] M.P. Washburn, R. Ulaszek, C. Deciu, D.M. Schieltz, J.R. Yates 3rd, Anal. Chem. 74 (2002) 1650.

[37] J.R. Yates 3rd, E. Carmack, L. Hays, A.J. Link, J.K. Eng, Methods Mol. Biol. 112 (1999) 553.
[38] M.P. Washburn, D. Wolters, J.R. Yates 3rd, Nat. Biotechnol. 19 (2001) 242.
[39] W.H. McDonald, R. Ohi, D.T. Miyamoto, T.J. Mitchison, I. Yates, R. John, Int. J. Mass Spectrom. 219 (2002) 245.
[40] M. Dreger, L. Bengtsson, T. Schoneberg, H. Otto, F. Hucho, Proc. Natl. Acad. Sci. USA 98 (2001) 11943.
[41] T. Hermann, M. Finkemeier, W. Pfefferle, G. Wersch, R. Kramer, A. Burkovski, Electrophoresis 21 (2000) 654.
[42] P.D. von Haller, S. Donohoe, D.R. Goodlett, R. Aebersold, J.D. Watts, Proteomics 1 (2001) 1010.
[43] V. Soskic, J. Godovac-Zimmermann, Proteomics 1 (2001) 1364.
[44] M. Ferro, D. Seigneurin-Berny, N. Rolland, A. Chapel, D. Salvi, J. Garin, J. Joyard, Electrophoresis 21 (2000) 3517.
[45] S. Ohlmeier, C. Scharf, M. Hecker, Electrophoresis 21 (2000) 3701.
[46] T.P. Conrads, G.A. Anderson, T.D. Veenstra, L. Pasa-Tolic, R.D. Smith, Anal. Chem. 72 (2000) 3349.
[47] Y. Shen, N. Tolic, R. Zhao, L. Pasa-Tolic, L. Li, S.J. Berger, R. Harkewicz, G.A. Anderson, M.E. Belov, R.D. Smith, Anal. Chem. 73 (2001) 3011.
[48] R.D. Smith, G.A. Anderson, M.S. Lipton, C. Masselon, L. Pasa-Tolic, Y. Shen, H.R. Udseth, Proteomics 6 (2002) 61.
[49] L. Pasa-Tolic, R. Harkewicz, G.A. Anderson, N. Tolic, Y. Shen, R. Zhao, B. Thrall, C. Masselon, R.D. Smith, J. Am. Soc. Mass Spectrom. 13 (2002) 954.
[50] R.D. Smith, G.A. Anderson, M.S. Lipton, L. Pasa-Tolic, Y. Shen, T.P. Conrads, T.D. Veenstra, H.R. Udseth, Proteomics 2 (2002) 513.
[51] M.S. Lipton, L. Pasa-Tolic, G.A. Anderson, D.J. Anderson, D.L. Auberry, J.R. Battista, M.J. Daly, J. Fredrickson, K.K. Hixson, H. Kostandarithes, C. Masselon, L.M. Markillie, R.J. Moore, M.F. Romine, Y. Shen, E. Stritmatter, N. Tolic, H.R. Udseth, A. Venkateswaran, K.K. Wong, R. Zhao, R.D. Smith, Proc. Natl. Acad. Sci. USA 99 (2002) 11049.
[52] M. Palmblad, M. Wetterhall, K. Markides, P. Hakansson, J. Bergquist, Rapid Commun. Mass Spectrom. 14 (2000) 1029.
[53] R. Tonge, J. Shaw, B. Middleton, R. Rowlinson, S. Rayner, J. Young, F. Pognan, E. Hawkins, I. Currie, M. Davison, Proteomics 1 (2001) 377.
[54] D. Chelius, P.V. Bondarenko, J. Proteome Res. 1 (2002) 317.
[55] O. Stemmann, H. Zou, S.A. Gerber, S.P. Gygi, M.W. Kirschner, Cell 107 (2001) 715.
[56] M. Geng, J. Ji, F.E. Regnier, J. Chromatogr. A 870 (2000) 295.
[57] J. Ji, A. Chakraborty, M. Geng, X. Zhang, A. Amini, M. Bina, F. Regnier, J. Chromatogr. B Biomed. Sci. Appl. 745 (2000) 197.
[58] M. Munchbach, M. Quadroni, G. Miotto, P. James, Anal. Chem. 72 (2000) 4047.
[59] M. Schnolzer, P. Jedrzejewski, W.D. Lehmann, Electrophoresis 17 (1996) 945.
[60] X. Yao, A. Freas, J. Ramirez, P.A. Demirev, C. Fenselau, Anal. Chem. 73 (2001) 2836.
[61] O.A. Mirgorodskaya, Y.P. Kozmin, M.I. Titov, R. Korner, C.P. Sonksen, P. Roepstorff, Rapid Commun. Mass Spectrom. 14 (2000) 1226.
[62] Y.K. Wang, Z. Ma, D.F. Quinn, E.W. Fu, Anal. Chem. 73 (2001) 3742.
[63] I.I. Stewart, T. Thomson, D. Figeys, Rapid Commun. Mass Spectrom. 15 (2001) 2456.
[64] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[65] D.K. Han, J. Eng, H. Zhou, R. Aebersold, Nat. Biotechnol. 19 (2001) 946.
[66] T.J. Griffin, D.K. Han, S.P. Gygi, B. Rist, H. Lee, R. Aebersold, K.C. Parker, J. Am. Soc. Mass Spectrom. 12 (2001) 1238.
[67] H. Zhou, J.A. Ranish, J.D. Watts, R. Aebersold, Nat. Biotechnol. 20 (2002) 512.
[68] M.B. Smolka, H. Zhou, S. Purkayastha, R. Aebersold, Anal. Biochem. 297 (2001) 25.
[69] M. Smolka, H. Zhou, R. Aebersold, Mol. Cell Proteomics 1 (2002) 19.
[70] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. USA 96 (1999) 6591.
[71] T.P. Conrads, K. Alving, T.D. Veenstra, M.E. Belov, G.A. Anderson, D.J. Anderson, M.S. Lipton, L. Pasa-Tolic, H.R. Udseth, W.B. Chrisler, B.D. Thrall, R.D. Smith, Anal. Chem. 73 (2001) 2132.
[72] L. Pasa-Tolic, P.K. Jensen, G.A. Anderson, M.S. Lipton, K.K. Peden, S. Martinovic, N. Tolic, J.E. Bruce, R.D. Smith, J. Am. Chem. Soc. 121 (1999) 7949.
[73] S.E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, M. Mann, Mol. Cell Proteomics 1 (2002) 376.
[74] J. Krijgsveld, R.F. Ketting, J. Johansen, R.H.A. Plasterk, A.J.R. Heck, submitted.
[75] J.M. Pratt, D.H. Robertson, S.J. Gaskell, I. Riba-Garcia, S.J. Hubbard, K. Sidhu, S.G. Oliver, P. Butler, A. Hayes, J. Petty, R.J. Beynon, Proteomics 2 (2002) 157.
[76] R. Zhang, C.S. Sioma, S. Wang, F.E. Regnier, Anal. Chem. 73 (2001) 5142.
[77] R. Zhang, C.S. Sioma, R.A. Thompson, L. Xiong, F.E. Regnier, Anal. Chem. 74 (2002) 3662.
[78] R. Zhang, F.E. Regnier, J. Proteome Res. 1 (2002) 139.
[79] J.S. Choudhary, W.P. Blackstock, D.M. Creasy, J.S. Cottrell, Trends Biotechnol. 19 (2001) S17.
[80] A. Pandey, M. Mann, Nature 405 (2000) 837.
[81] R. Aebersold, D.R. Goodlett, Chem. Rev. 101 (2001) 269.
[82] C. O'Donovan, R. Apweiler, A. Bairoch, Trends Biotechnol. 19 (2001) 178.
[83] A. Castagna, N. Campostrini, A. Farinazzo, G. Zanusso, S. Monaco, P.G. Righetti, Electrophoresis 23 (2002) 339.
[84] M.J. Hubbard, P. Cohen, Trends Biochem. Sci. 18 (1993) 172.
[85] P. Cohen, Trends Biochem. Sci. 17 (1992) 408.
[86] T. Hunter, Methods Enzymol. 200 (1991) 3.
[87] B.M. Sefton, Methods Enzymol. 201 (1991) 245.
[88] T. Hunter, Cell 80 (1995) 225.
[89] G.G. Mason, R.Z. Murray, D. Pappin, A.J. Rivett, FEBS Lett. 430 (1998) 269.

[90] K.L. Bennett, A. Stensballe, A.V. Podtelejnikov, M. Moniatte, O.N. Jensen, J. Mass Spectrom. 37 (2002) 179.
[91] M.P. Molloy, P.C. Andrews, Anal. Chem. 73 (2001) 5387.
[92] M.R. Larsen, G.L. Sorensen, S.J. Fey, P.M. Larsen, P. Roepstorff, Proteomics 1 (2001) 223.
[93] A. Stensballe, S. Andersen, O.N. Jensen, Proteomics 1 (2001) 207.
[94] T.P. Conrads, H.J. Issaq, T.D. Veenstra, Biochem. Biophys. Res. Commun. 290 (2002) 885.
[95] G. Neubauer, M. Mann, Anal. Chem. 71 (1999) 235.
[96] H. Steen, B. Kuster, M. Fernandez, A. Pandey, M. Mann, Anal. Chem. 73 (2001) 1440.
[97] M.B. Goshe, T.D. Veenstra, E.A. Panisko, T.P. Conrads, N.H. Angell, R.D. Smith, Anal. Chem. 74 (2002) 607.
[98] Y. Oda, T. Nagasu, B.T. Chait, Nat. Biotechnol. 19 (2001) 379.
[99] H. Zhou, J.D. Watts, R. Aebersold, Nat. Biotechnol. 19 (2001) 375.
[100] M.B. Goshe, T.P. Conrads, E.A. Panisko, N.H. Angell, T.D. Veenstra, R.D. Smith, Anal. Chem. 73 (2001) 2578.
[101] S.B. Ficarro, M.L. McCleland, P.T. Stukenberg, D.J. Burke, M.M. Ross, J. Shabanowitz, D.F. Hunt, F.M. White, Nat. Biotechnol. 20 (2002) 301.
[102] M.C. Posewitz, P. Tempst, Anal. Chem. 71 (1999) 2883.
[103] P. Cao, J.T. Stults, Rapid. Commun. Mass Spectrom. 14 (2000) 1600.
[104] D.C. Neville, C.R. Rozanas, E.M. Price, D.B. Gruis, A.S. Verkman, R.R. Townsend, Protein Sci. 6 (1997) 2436.
[105] W. Xhou, B.A. Merrick, M.G. Khaledi, K.B. Tomer, J. Am. Soc. Mass Spectrom. 11 (2000) 273.
[106] J. Hirabayashi, K. Kasai, Trends Glycosci. Glycotechnol. 12 (2000) 1.
[107] R.A. Laine, Glycobiology 4 (1994) 759.
[108] J. Hirabayashi, K. Kasai, J. Chromatogr. B 771 (2002) 67.
[109] P.A. Haynes, R. Aebersold, Anal. Chem. 72 (2000) 5402.
[110] S.M. Gomez, J.N. Nishio, K.F. Faull, J.P. Whitelegge, Mol. Cell Proteomics 1 (2002) 46.
[111] X.D. Wu, R.A. Kircher, P.H. McVerry, D.A. Malinzak, Dev. Biol. (Basel) 103 (2000) 61.
[112] G. Vecchione, B. Casetta, R. Santacroce, M. Margaglione, Rapid Commun. Mass Spectrom. 15 (2001) 1383.