Ann Ist Super Sanità 2005;41(4):443-450

Mass spectrometry for protein identification and the study of post translational modifications

Anna Maria SALZANO and Marco CRESCENZI

Dipartimento di Ambiente e Connessa Prevenzione Primaria, Istituto Superiore di Sanità, Rome, Italy

Summary. - Mass spectrometry is a powerful tool for the structural characterization of proteins. The development of a large number of analytical strategies based on different mass spectrometric techniques has established mass spectrometry as an indispensable technology to interpret the information encoded in the genome. This manuscript intends to provide a rapid view of principles and instrumentation in mass spectrometry-based proteomics, together with a description of the basic features and limits of the analysis of post translational modifications by mass spectrometry. The application of one of the available mass spectrometry technologies for protein identification to a multi-protein complex isolated in a specific functional context is reported. The identification of a phosphorylation site on a single, purified protein is also illustrated as an example.

Key words: proteomics, mass spectrometry, post translational modifications, phosphorylation site identification.

Riassunto (*Uso della spettrometria di massa per l'identificazione di proteine e lo studio di modificazioni post-traduzionali*). - La spettrometria di massa è un potente mezzo per la caratterizzazione strutturale di proteine. Lo sviluppo di un elevato numero di strategie analitiche, basate sull'impiego di differenti tecniche spettrometriche, ha reso la spettrometria di massa una tecnologia indispensabile per l'interpretazione delle informazioni codificate nel genoma. Questo manoscritto vuole fornire una rapida visione di insieme dei principi e delle strumentazioni utilizzate per l'analisi proteomica, oltre ad una descrizione dei fondamenti e dei limiti dell'analisi delle modifiche post-traduzionali mediante spettrometria di massa. È inoltre riportato un esempio di applicazione di una delle tecnologie di spettrometria di massa per la identificazione di proteine ad un complesso multiproteico isolato in uno specifico contesto funzionale. L'identificatione di un sito di fosforilazione su una proteina purificata è stato inoltre illustrato come esempio.

Parole chiave: proteomica, spettrometria di massa, modifiche post traduzionali, identificazione di siti di fosforilazione.

Introduction

Mass spectrometry (MS) has been used for the analysis of proteins and peptides since 1989, when two new "soft" techniques for gas phase ionization of large, polar, and highly charged molecules were introduced [1, 2]. In one of these, electrospray ionization (ESI), ions are formed from a liquid solution at atmospheric pressure, while in the second, matrix assisted laser desorption ionization (MALDI), a laser pulse induces the sample to sublimate out of a dry crystalline matrix. The devices that bring the analytes into gas phase and ionize them are known as "sources".

There are four basic types of mass analyzers used for protein studies: the ion trap (IT), time of flight (TOF), quadrupole (Q), and Fourier transform ion cyclotron (FT-ICR) devices. These are very different in design and performance, each with its own strengths and weaknesses. Two analysers can be placed in tandem to perform two-stage mass spectrometry (tandem MS, commonly referred to as MS/MS). In MS/MS, with diverse mechanisms in different mass spectrometers, peptide ions are isolated and fragmented and the mass-to-charge ratio of the fragments is measured. The most common tandem mass analyzers, able to perform peptide fragmentation,

Indirizzo per la corrispondenza (Address for correspondence): Marco Crescenzi, Dipartimento di Ambiente e Connessa Prevenzione Primaria, Istituto Superiore di Sanità, V.le Regina Elena 299, 00161 Roma. E-mail: crescenz@iss.it.

are quadrupole ion trap (IT), triple quadrupole (QqQ), quadrupole-TOF (Q-TOF) and TOF-TOF. The performance of different tandem mass spectrometers varies in terms of mass accuracy, mass resolution, robustness, and ease of operation.

Different analyzers can be coupled with various sources, so that the number of possible instrumental configurations is large. In particular, two different mass spectrometer configurations were used to produce most of the proteomic data published thus far: ion traps coupled to ESI sources (ESI-IT) and TOFs coupled to MALDI sources (MALDI-TOF).

ESI-IT instruments are very robust, comparatively inexpensive, and easy to operate, although their accuracy is relatively low. A recent development of the ion trap analyser is the "linear" or "two-dimensional" ion trap that overcomes the accuracy limit of traditional ion traps (three-dimensional trap), achieving also an increase in sensitivity and resolution [3, 4].

MALDI-TOF instruments have the advantages of greater sensitivity and higher resolution power, although they cannot produce sequence information. More recent instruments have a MALDI source coupled to Q-TOF [5] or TOF-TOF [6] analyzers and are therefore able to perform peptide sequencing taking advantage of the sensitivity of MALDI sources.

Instrument configurations including linear ion trap and FT-ICR mass spectrometers are likely to become widespread in proteomics laboratories in the very near future.

Protein identification by mass spectrometry

MS experiments for protein identification are mainly based on peptide analysis. Mass spectrometry can measure the mass of intact proteins, but the sensitivity of the mass spectrometer for proteins is much lower than for peptides; moreover, the mass of whole proteins cannot be measured accurately enough for identification purposes; finally, proteins are difficult to handle and most solubilizing conditions are not compatible with MS.

Nevertheless, some sequence information is required for protein identification in most cases. MS is much more suitable to obtain sequence information from peptides (up to 20 amino acids) than intact proteins. Partial sequence information can be obtained from intact proteins with very sophisticated instruments, such as FT-ICR mass spectrometers, through a recent approach called "top-down" [7-10].

In the first step of a typical proteomic experiment, the proteins of interest are purified from cells or tissues using different methods, such as physical and/or biochemical fractionation, the latter essentially based on chromatographic techniques, including affinity purification. This is actually the most critical and challenging stage, because of the importance of working with a biologically significant subset of proteins and the difficulty of obtaining it.

The partially purified protein sample is generally further fractionated through mono-dimensional gel electrophoresis (1D-gel). These strategies concern above all functional proteomic studies.

Whole cell and tissue proteins can also be directly separated by bi-dimensional polyacrylamide gel electrophoresis (2D-PAGE), a powerful method based on orthogonal separation of proteins by isoelectric point and molecular weight [11, 12]. 2D-PAGE is still one of the most used technologies for comprehensive proteomic studies.

The gel band (1D gel) or spot (2D gel) is excised and the proteins are in gel-digested using different proteases and/or chemicals. Trypsin is a very stable and efficient protease that specifically cleaves at the C-terminal side of lysine and arginine residues. Peptides generated by trypsin cleavage have a basic residue at the carboxy-terminus and, on average, the right size to be usefully detected and sequenced in a mass spectrometer. Therefore, trypsin is most commonly used in identification studies.

In addition to the above mentioned strategies, in which protein mixtures are first separated on 1D or 2D gel electrophoresis and then further processed for MS analysis, many other methods have been reported. Widely used "gel-free" methods mainly rely on the bidimensional fractionation of peptide mixtures. The protein sample (an unfractionated cellular lysate or a partially purified protein mixture) is reduced, alkylated, and digested in solution, then the resulting peptide mixture is fractionated through a strong cationic exchange column (SCX), which separates peptides on the base of their charge, and the derived fractions are further separated on a reverse phase (RP) column. This "multidimentional chromatography" step is used to fractionate highly complex protein mixtures in the technique known as "multidimensional protein identification technology" (MudPIT) [13]. Currently, this approach appears to be the best choice to analyze samples that cannot be efficiently resolved on gels owing to their physico-chemical properties.

In most cases, the peptide mixtures generated by digesting multiple protein mixtures are not introduced directly into the mass spectrometer but are first separated by high pressure liquid chromatography (HPLC). Peptides are eluted from RP columns using increasing concentrations of organic solvents, which allow the separation of peptides on the basis of their different hydrophobicity. The chromatographic column is coupled on-line with a mass spectrometer equipped with an ESI source that ionizes samples from liquid phase, in the so-called LC-ESI-MS techniques.

In LC-ESI-MS, signal intensity is directly proportional to analyte concentration, so peptides should be eluted in as small a volume as possible. This can be achieved by reducing the internal diameters of chromatographic columns, as much as compatible with their remaining clog-free and capable of being efficiently packaged; the most useful capillary columns have 50-150 μm inner diameters and can work at flow rates as low as 100 nl/min. Such low flows can produce a stable spray in specially designed electrospray sources, called nanospray sources. Instruments so devised are known as nano-LC-MS systems.

Once the peptide mixture has been analyzed by MS, proteins are identified by matching a list of experimental peptide masses, measured with high accuracy, with the theoretically calculated peptide masses obtained from an *in silico* digestion of all proteins present in a given database, considering the specificity of the protease employed. This approach is called "peptide mass fingerprinting" (PMF). For PMF analysis several search algorithms are currently available. PMF has largely been used for identification of protein spots purified from 2D gel and analysed by MALDI-TOF.

However, protein identification is more efficiently attained by using sequence information, derived from collision induced fragmentation ion spectra (CID spectra), obtained with instruments able to perform MS/MS analyses. In MS/MS experiments, a specific peptide ion is isolated and its fragmentation is induced by collision with an inert gas. Fragmentation "rules" for peptides exist [14] that help in the interpretation of tandem mass spectra; a common nomenclature of sequence ions in MS/MS spectra has been proposed [15].

Fragmentation spectra cannot be used to derive a full peptide sequence, i.e. the sequence of a peptide cannot be determined *de novo* in a MS experiment, but they are matched to comprehensive protein databases, by using different algorithms such as Peptide Sequence Tags [16], Sequest [17], or Mascot [18]. The output of such comparisons is the identity of peptides that subsequently are used to compile a protein hit list. Usually a protein can be confidently identified by two or three peptides. If only a single peptide identifies a protein, further confirmation is needed. For example, a peptide with the same sequence can be synthesised and its chromatographic properties and fragmentation spectrum can be compared with the peptide found in the identification experiment.

Global proteomic analysis of a DNA-replication complex from Hela cells

A protein identification study has been carried out in our laboratory in order to characterize the protein composition of a DNA-replication complex from eukaryotic cells. Many of the constituents of the DNA replication apparatus, in fact, associate with other cellular factors as components of multiprotein complexes, which act cooperatively in networks to regulate cell cycle progression and checkpoint control. Nuclear extracts obtained from Hela cell have been progressively fractionated through gel filtration and anionic exchange (Mono S) columns while monitoring the replication complex with a specific DNA-polymerase assay [19]. Finally, the fractions demonstrating DNA polymerase activity have been further purified by affinity chromatography on a heparin column. The resulting protein sample has been analyzed by mono-dimensional gel electrophoresis and stained to visualize proteins. As expected, the purified material was still very complex. Thus, the entire gel lane containing the replication complex has been subdivided into 20 slices, each of which has been subjected to reduction, alkylation, and in gel trypsinization as described above. Peptide mixtures were analyzed by LC-ESI-MS/MS, using a micro-HPLC coupled to an ion trap mass spectrometer equipped with a microspray source. Chromatographic and MS conditions have been specifically set up to take into account sample complexity. The acquisition method for LC-ESI-MS/MS experiment has been set to perform MS/MS data-dependent scanning of the three most abundant ions, enabling the dynamic exclusion function in order to acquire as many fragmentation spectra as possible.

MS/MS spectra were used to search human entries from both the NCBI database and the UniProt Knowledgebase database (UniProtKB). UniProtKB is the central access point for extensive curated protein information available at http://www.expasy.uniprot.org/. Both the Sequest and the Mascot algorithms were used for protein identification. Results from different searches were manually inspected, filtered according to matching scores, and finally compared in order to derive a unique list of identified proteins.

The experiment was repeated, starting from an independently purified protein complex, to evaluate the biological variability of the samples. Altogether, 232 proteins were identified. In order to group the identified proteins in subsets sharing common functional or biological characteristics, the UniProt database was queried. Functional annotations were integrated with other annotations covering different biological aspects. A schematic representation is shown in Fig. 1.

To attain an in-depth biological analysis of the identified protein set and eventually detect relevant subsets, a program named Pandora (Protein Annotation Diagram Oriented Analysis), available at http://www.pandora.cs.huji.ac.il/, was employed [20].

Inspection of the protein list revealed, together with the presence of proteins involved in DNA replication, also proteins known to act in DNA repair. Intriguingly,

all six minichromosome maintenance (MCM) proteins were also identified. However, such a large identification study requires that the participation of each candidate protein to DNA replication complexes be confirmed by independent means. Indeed, it is noteworthy that not only proteins known to effect or plausibly involved in DNA replication were found. In fact, a significant contamination from proteins participating in RNA splicing/maturation was evident. This is not surprising because of the difficulty of obtaining a homogeneous protein complex only through chromatographic purification. Data obtained from this proteomic study will be used as a starting point to investigate the intricacies of the DNA replication machinery using other biochemical approaches.

Post translational modifications analysis

Post translational modifications (PTMs) are covalent modifications that regulate protein functions, determining their activity state, cellular location and dynamic interactions with other proteins. More than 300 different types of PTMs are known, and new ones are regularly discovered. Despite the importance of PTMs to understand a great deal of cellular processes, their comprehensive analysis has encountered some biological and analytical limitations. PTMs are often transient and occur in vivo only in a small fraction of proteins (less than 1%), principally because they are present in substoichiometric amounts on protein molecules. Therefore, the isolation of a sufficiently large amount of modified proteins for biochemical studies is not easily attained and, almost invariably, the sample consists of a heterogeneous mixture of modified and unmodified proteins.

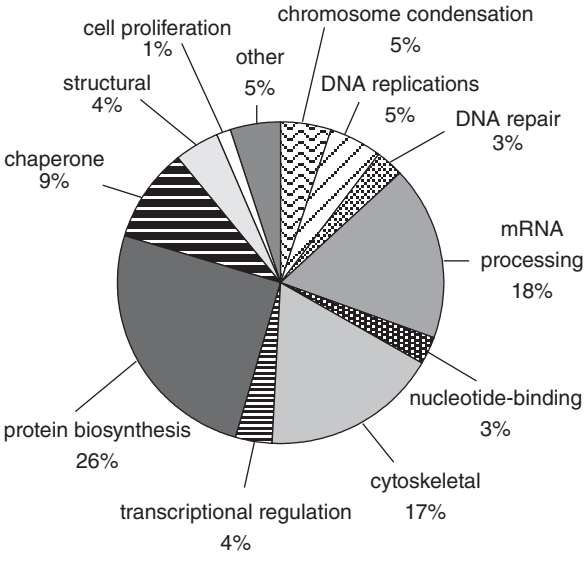

Fig. 1. - Graphic representation of the identified proteins, grouped on the basis of their functional role and the processes they are involved in. Annotations for each protein were obtained from the UniProt database.

The most important and best studied PTMs include reversible phosphorylation, occurring at threonine, serine and tyrosine residues, and glycosylation, consisting of the covalent attachment of olygosaccarides to asparagine (so called N-linked) or serine/threonine (O-linked) residues. Other common modifications are acetylation, methylation, lipid attachment, sulfation of tyrosine, ubiquitination and disulphide bond formation.

MS is a general method for modification analysis, because PTMs lead to a mass increase or decrease with respect to the molecular weight expected on the basis of protein sequence. Nowadays, MS is widely used to determine the type and the site of modifications on single, purified proteins. Once a protein has been isolated, a variety of techniques can be used to identify the modification and identify the modified residues. PTMs studies on intact proteins are very rare. In very few cases the molecular weight of the intact protein can be established by MS, for example, if the protein is sufficiently homogeneous and its molecular weight is less than 100 kDa. Moreover, this kind of investigation can be afforded only with sophisticated instruments, such as FT-ICR mass spectrometers, because of their exceptional sensitivity, mass resolution and mass accuracy.

Detailed characterization of modifications is obtained, almost invariably, subjecting the protein to enzymatic or chemical degradation. PTM analysis requires the coverage of as much of the protein sequence as possible; for this purpose, multiple enzymes with different cleavage specificities must be employed. Proteolytic mixtures are then analysed by MS, primarily to determine molecular masses. The experimentally determined peptide masses are matched to the list of expected masses, as computed from the protein sequence. Once all possible non-specific cleavages or contaminants have been taken into account, unmatched signals are inspected for mass differences corresponding to known modifications. A comprehensive list of mass differences corresponding to PTMs (Δmass values) can be found at http://abrf.org/index.cfm/dm.home.

Some computer programs can be helpful in interpreting PMF experiments aimed at PTM identification; among them, GPMAW, a proprietary software, is one of the most powerful tools available today.

In most cases peptide mass measurements are not sufficient to determine the nature of the modification

and, for this reason, peptides are fragmented by tandem MS. In mass spectrometers used for proteomic studies, fragmentation occurs principally at peptide bonds, while modified aminoacids remain intact so that the fragmentation pattern is similar to the unmodified peptide, and the differences derive from the presence of a residue carrying a mass increment.

The identification of modification sites can be obtained through manual inspection of fragmentation spectra of modified peptides, if the identity and sequence of the protein are already known. When the modification is labile it could be lost during or even before fragmentation. The loss of the modifying group is diagnostic of the presence of the modification on the intact peptide. Several MS-based analytical strategies taking advantage of this phenomenon have been improved in the last few years. In particular, the "neutral loss" [21, 22] and "precursor ion scanning" [23, 24] techniques allow highly specific and sensitive detection of modified peptides. However, the analysis of the modification state of a purified protein remains a challenging analytical endeavour.

The amount of purified protein needed for PTM analysis is much higher than that required for protein identification, because of the need to cover almost the entire sequence. A useful recommendation to biologists who intend to take on a PTM study is to purify at least 1 μg of protein. Nevertheless, the development of new technologies is rapidly accelerating the laborious task of PTM analysis. In particular, encouraging results have been obtained since the introduction of a new generation of FT-ICR-MS instruments, in part owing to their ability to be coupled with electron capture dissociation (ECD), a relatively new technology for peptide and protein fragmentation with great potential for PTM identification [25].

Recently, attempts have been made to define modifications on a proteome-wide scale. Given the difficulties of identifying all modifications even on a single protein, it is clear that, at present, scanning for proteome-wide modifications can not be comprehensive. However, a large amount of biologically useful information can be generated by this approach. Programs such as Sequest and Mascot have recently been improved to extend searching capabilities to modified peptides. One or more modifications can be included in the search, although this carries the cost of a vast increase in search space and a corresponding decrease in the confidence of identification. For this reason, one of the strategies used is an extension of the approach employed for the analysis of complex protein mixtures: the experiment is usually divided into identification of a set of proteins from non-modified peptides, followed by searching only those proteins for modified peptides [26].

A more functionally oriented approach focuses on the search for one type of modification on all proteins present in a sample. This strategy requires an affinity-based enrichment of post translationally modified proteins and peptides to increase the relative abundance of a selected class of modified polypeptides. Anti-phosphotyrosine and anti-phosphoserine/phosphothreonine antibodies have been usefully employed to selectively precipitate phosphorylated proteins to study cell signalling [27, 28].

Immobilized metal affinity chromatography (IMAC) has been widely used for purification of phosphorylated peptides from relatively simple mixtures before MS analysis. It exploits the affinity of Fe(III) and Ga(III) ions toward phosphate groups. In global phosphoproteomic experiments, the IMAC method has been refined introducing, before affinity purification, O-methylesterification of carboxylic groups to reduce non specific binding [29], or a strong anion exchange chromatography step to reduce sample complexity [30]. Several chemical approaches for introducing affinity tags to selectively purify modified peptides seem very promising. One of the tags most often used for these purposes is biotin. Enrichment of phosphopeptides and phosphoproteins has been obtained through β-elimination of phosphoric acid from phosphoserines and phosphothreonines and subsequent Michael-addition reaction to attach the affinity tag [31, 32].

Affinity tagging has been successfully adopted to purify nitrotyrosine-containing proteins: the nitrogroup has been reduced to nitrosotyrosine and then this group has been biotinylated for affinity enrichment [33].

PTM analysis is important for the understanding of all physiological processes, especially of eukaryotic organisms. The presence of a given PTM cannot be predicted, with a high degree of confidence, by computational sequence analysis and therefore for their determination experimental proteomic techniques have to be developed. Today, a large number of proteomics and mass spectrometry laboratories are involved in research projects focused on the characterization of PTMs characterizing a specific protein in a particular physiological and/or pathological state. Very often during routine investigations, a clear indication of the presence of a PTM on a given protein comes, sometimes unexpectedly, from 2D-PAGE analysis. 2D-PAGE is, in fact, a high resolution technique, able to separate different modified forms of the same protein into different spots on a gel [34]. Different forms of phosphorylated proteins can be visualized on a 2D-PAGE gel through phosphorimaging of proteins radiolabeled with 32P [35] or staining with a specific dye such as PRO Q diamond [36].

Analysis of phosphorylation of an affinity-purified protein

An interesting example of PTM analysis by MS, carried out in our laboratory, was the characterization of PTMs on a protein, purified from an ancient eukaryotic organism, that was suspected to be modified because of its abnormal migration in 2D-PAGE. The protein was affinity purified and loaded onto a monodimensional, 4-12%-polyacrylamide, SDS gel, and stained with Comassie. The protein's migration properties were compared with those of the recombinant protein expressed in bacteria, hence presumably not modified. The native protein migrated more slowly than its recombinant counterpart and, moreover, produced a broader band. Excised gel bands were reduced and alkylated essentially as described [37] and digested with trypsin. MALDI-TOF spectra of the native protein tryptic digest were inspected in a search for post translational modifications, by directly comparing them with MALDI-TOF spectra of the recombinant sample. This comparison unveiled the presence in the native sample (Fig. 2, panel B) of a signal shifted up by 80 Da compared to that found in the recombinant protein spectrum (Fig. 2, panel A). Such shift could originate from the phosphorylation of the peptide corresponding to the mass measured in the recombinant protein spectrum. In order to confirm this hypothesis, the peptide mixtures were analysed by MALDI-TOF-MS in both linear and reflectron modes. MALDI-TOF-MS analysis of phosphopeptides is a key step for their recognition. In MALDI-TOF mass spectrometers, loss of HPO_3 and/or H_3PO_4 occurs post source as a result of metastable decomposition (post source decay, PSD), giving rise to fragments having the same velocity as their parent ions but with reduced energy. Fragment ions cannot be distinguished from their parents if the spectra are collected in linear mode, but are detected at lower apparent masses in reflectron mode [38]. In our case, a signal shifted down by 98 Da, corresponding to the loss of H_3PO_4 from the putative phosphopeptide, became visible in reflectron mode (Fig. 2, panel B). This fragmentation is common for serine- and threonine-phosphorylated peptides, while it is not favoured in the case of phosphotyrosine.

In addition, we acquired MALDI-TOF spectra in reflectron mode using a specific matrix preparation, 2,5-dihydroxybenzoic acid (DHB) in 1% ortho-phosphoric acid, which has recently been reported to enhance phosphopetide detection [39]. The metastable decomposition product of our phosphopeptide was no longer detectable in the spectrum obtained with this matrix (Fig. 2, panel C).

Further characterization of the phosphopeptide was obtained by LC-ESI-MS/MS analysis of tryptic digests

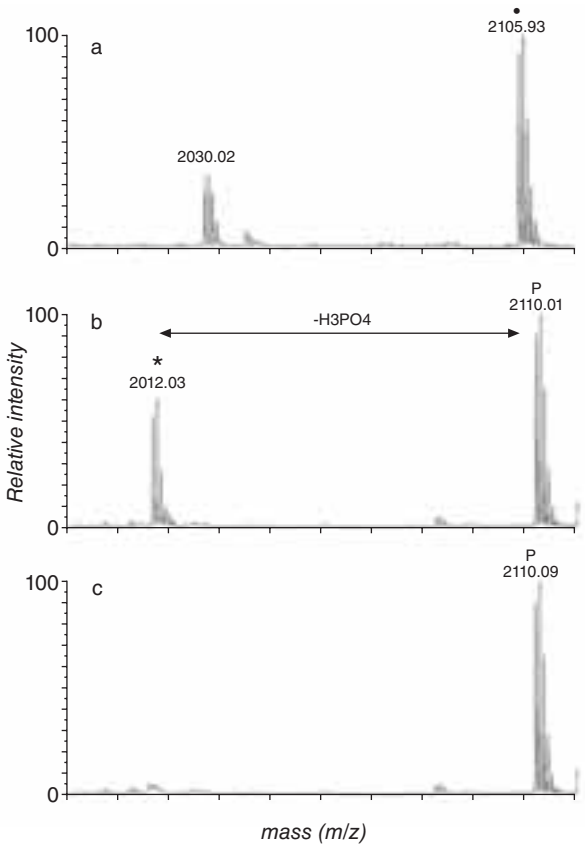

Fig. 2. - Phosphopeptide identification by MALDI-TOF mapping. **Panel A** shows the MALDI-TOF reflectron spectrum of the tryptic digest of the recombinant protein. The unphosphorylated peptide signal is at m/z 2030.02. The peptide indicated with • is not related to this phosphorylation analysis and is only present in the recombinant sample. **Panel B** shows the MALDI-TOF reflectron spectrum of the native protein acquired using alpha cyano-4-hydroxycinnamic acid as matrix. The peak at m/z 2110.02, indicated with "P" corresponds to the phosphorylated peptide; the peak indicated with * corresponds to the metastable ion derived from the loss of phosphoric acid. **Panel C** shows the MALDI-TOF reflectron spectrum of the native protein acquired using DHB in 1% ortho-phosphoric acid in which the metastable peak is no longer evident.

of the native protein using a microcapillary HPLC system and an ion trap mass spectrometer equipped with an ESI source. Collision induced dissociation (CID) of phosphopeptides performed with IT instruments gives rise in the majority of cases to trace fragments that can be easily detected [40]. MS/MS spectra obtained by LC-ESI-IT-MS/MS analysis were inspected for the loss of 98, 49, 32 m/z units from parent ions, corresponding to the loss of one phosphate group for singly, doubly and triply charged parent ions. Fragmentation of the species at m/z 1054.6,

corresponding to the doubly charged ion of the putative phosphorylated peptide, gave a strong signal at m/z 1006.4, with a neutral loss of 49 Da. The identification of the phosphorylation site was then obtained through an LC-MS^3 analysis in which the ion at m/z 1006.4 was isolated and further fragmented. Interpretation of the MS^3 spectrum lead to the identification of the formerly phosphorylated residue, recognized by the presence of a dehydrobutyric acid in place of a threonine residue, as previously reported for peptides undergoing loss of H_3PO_4 [41].

The analytical strategy adopted in this study is of general applicability for identifying phosphorylation sites with MALDI-TOF and ESI-IT instruments.

Acknowledgements

We thank G. Maga, CNR, Pavia, and E. Pozio and M. Lalle, Istituto Superiore di Sanità, Roma, for allowing us to use as examples in this paper data generated in collaborative studies with them.

Received on 5 September 2005.
Accepted on 13 October 2005.

REFERENCES

1. Fenn J B, Mann M, Meng C K, Wong S F, Whitehouse C M. Electrospray ionization for the mass spectrometry of large biomolecules. *Science* 1989;246:64-71.

2. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular mass exceeding 10000 daltons. *Anal Chem* 1989;60:2299-301.

3. Hager J W. A new linear ion trap mass spectrometer. *Rapid Comm Mass Spectrom* 2002;16:512-26.

4. Shwartz J C, Senko M W, Syka J E. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* 2002;13:656-69.

5. Krutchinsky A N, Kalkun M, Chait B T. Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. *Anal Chem* 2001;73:5066-77.

6. Medzihradszky K F *et al*. The characteristic of peptide collision-induced dissociation using a high performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem* 2000;72:552-8.

7. Mortz E *et al*. Sequence tag identification of intact proteins by matching tandem mass spectral data against sequence databases. *Proc Natl Acad Sci USA* 1996;93:8264-67.

8. Horn D M, Zubarev R A, McLafferty. F W. *Proc Natl Acad Sci USA* 2000;97:10313-7.

9. Sze S K, Ge Y, Oh H, McLafferty F W. Top-down mass spectrometry of a 29-KDa protein for characterization of any post translational modification to within one residue. *Proc Natl Acad Sci USA* 2002;99:1774-9.

10. Taylor G K *et al*. Web and database software for identification of intact proteins using top-down mass spectrometry. *Anal Chem* 2003;75:4081-6.

11. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins *J Biol Chem* 1975;250(10):4007-21.

12. Rabilloud T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climb up the mountain. *Proteomics* 2002;2:3-10.

13. Washburn MP, Wolters D, Yates JR. 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001;19(3):242-7.

14. Biemann K. Mass spectrometry of peptides and proteins. *Annu Rev Biochem* 1992;61:977-1010.

15. Roepstorff P, Fohmann J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 1984;11:601.

16. Mann M, Wilm M. Error tolerant identification of peptides in sequence database by peptide sequence tags. *Anal Chem* 1994;66:4390-9.

17. Eng J K, McCormack A L, Yates J R. An approach to correlate MS/MS data to amino acid sequences in proten database. *J Am Soc Mass Spectrom* 1994;5:976-89.

18. 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* 1999;20:3551-67.

19. Maga G, Hübscher U. DNA replication machinery: functional characterization of a complex containing DNA polymerase alpha, DNA polymerase delta, and replication factor C suggests an asymmetric DNA polymerase dimer. *Biochemistry* 1996;35:5764-77.

20. Kaplan N, Vaaknin A, Linial M. PANDORA: keyword-based analysis of protein sets by integration of annotation sources. *Nucleic Acids Research* 2003;31:5617-26.

21. Covey T R , Sushan B I, Bonner R, Shroder W, Hucho F. LC/MS and LC/MS/MS screening for the sites of post-translational modifications in proteins. In: Jörnvall H, Höög J O, & Gustavsson A M (Ed.). *Methods in Protein Sequence Analysis*. Basel: Birkhauser Verlag; 1991. p. 411-47.

22. Bateman R H *et al*. A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass specrtrometer for studying prote phosphorylation. *J Am Soc Mass Spectrom* 2002;13:792-803.

23. Wilm M, Neubauer G, Mann M. Parent ion scans of unseparated peptide mixtures. *Anal Chem* 1996;68:527-33.

24. Carr S A, Huddleston M J, Annan R S. Selective detection and sequencing of phosphopeptides at femtomole level by mass spectrometry. *Anal Biochem* 1996;239:180-92.

25. Zubarev R A *et al*. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* 2000;72:563-73.

26. MacCoss M J *et al.* Shotgun identification of protein complexes and lens tissue. *Proc Natl Acad Sci USA* 2002;99:7900-5.

27. Pandey A *et al*. Analysis of receptor signaling pathways by mass-spectrometry: identification of Vav-2 as substrate of the epidermal and platelet-derived growth factor receptors. *Proc Natl Acad Sci USA* 2000;97:179-84.

28. Gronborg M, Kristiansen TZ, Stensballe A, Andersen J S, Ohara O, Mann M, Jensen O N, Pandey A. A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phosphspecific antbodies: identification of a novel pr Frigg, as a protein kinase A substrate. *Mol Cell Proteomics* 2002;1:517-27.

29. Ficarro S B, McCleland M L, Stukenberg P T, Burke D J, Ross M M, Shabanowitz J, Hunt D F, White F M. Phosphoproteome analysis by mass spectrometry and its application to Saccaromyces cerevisiae. *Nat Biotechnol* 2002;20:301-5.

30. Nühse T S, Stensballe A, Jensen O N, Peck S C. Large scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 2003;2:1234-43.

31. Zhou H, Watts J D, Aebersold R. A systematic approach to the analysis of protein phosphoryation. *Nat Biotechnol* 2001;19: 375-8.

32. Oda Y, Nagasu T, Chait B T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* 2001;19:379-82.

33. Nikov G, Bhat V, Wishnok J S, Tannebaum S R. Analysis of nitrated proteins by nitrotyrosine-specific affinity probes and mass spectrometry. *Anal Biochem* 2003;320:214-22.

34. Giannazza E, Celentano F, Magenes S, Ettori C, Righetti P G. Formula for immobilized pH gradients including pH extremes. *Electrophoresis* 1989;10:806-8.

35. Bykova N V, Stensballe A, Egsgaard H, Jensen O N, Moller I M. Phoshorylation of formate dehydrogenase in potato tuber mitochondria. *J Biol Chem* 2003;278:26021-30.

36. Patton W F. Detection technologies in proteome analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* 2002;771:3-31.

37. Shevchenko S, Wilm M, Vorm O, Mann, M. *Anal. Chem.* 1996;68:850-8.

38. Annan R S, Carr S A. Phosphopeptide analysis by matrix-assisted laser desorption time-of-flight mass spectrometry. *Anal Chem* 1996;68:3413-21.

39. Kjellstrom S, Jensen O. Phosphoric acid as a matrix additive for MALDI MS analysis of phosphopeptides and phosphoproteins. *Anal Chem*. 2004;1(76):5109-17.

40. DeGnore J P, Qin J. Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J Am Soc Mass Spectrom* 1998;9:1175-88.

41. Tholey A, Reed J, Lehmann W D. Electrospray tandem mass spectrometric studies of phosphopeptides and phosphopeptide analogues. *J Mass Spectrom* 1999;34:117-23.