

Minireview

Functional genomics by mass spectrometry

Jens S. Andersen, Matthias Mann*

Protein Interaction Laboratory, University of Southern Denmark, DK 5230 Odense M, Denmark

Received 19 May 2000

Edited by Gianni Cesareni

Abstract Systematic analysis of the function of genes can take place at the oligonucleotide or protein level. The latter has the advantage of being closest to function, since it is proteins that perform most of the reactions necessary for the cell. For most protein based ('proteomic') approaches to gene function, mass spectrometry is the method of choice. Mass spectrometry can now identify proteins with very high sensitivity and medium to high throughput. New instrumentation for the analysis of the proteome has been developed including a MALDI hybrid quadrupole time of flight instrument which combines advantages of the mass fingerprinting and peptide sequencing methods for protein identification. New approaches include the isotopic labeling of proteins to obtain accurate quantitative data by mass spectrometry, methods to analyze peptides derived from crude protein mixtures and approaches to analyze large numbers of intact proteins by mass spectrometry directly. Examples from this laboratory illustrate biological problem solving by modern mass spectrometric techniques. These include the analysis of the structure and function of the nucleolus and the analysis of signaling complexes. © 2000 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

Key words: Proteomics; Protein–protein interaction; Protein identification; Genomics; Nucleolus; Signaling; Phosphorylation

1. Introduction

During the 80's and particularly the 90's biological mass spectrometry has become a powerful method for the characterization of biomolecules such as proteins and nucleic acids. The main driving forces for this development have been novel ionization techniques which could transfer the biomolecules from the liquid phase to the gas phase which make them amenable to measurement in the mass spectrometer. The two dominant methods are electrospray ionization [1] and matrix assisted laser desorption ionization [2,3] which by historical coincidence were introduced to large biomolecules at the same time. With these two methods, the 90's have seen the application of concepts such as peptide mapping, which could only be used on large amounts of sample before, to more realistic protein amounts today. The detection of posttranslational modifications also became much easier and more sensitive with these methods. In 1993 a number of groups [4–8]

published a method to correlate the mass spectrometric information contained in a peptide mass map with the information in the expanding protein sequence databases. Even though 'mass fingerprinting' needed several more years to become a practically viable method for protein identification, in retrospect the publication of these papers marked the turning point at which mass spectrometry became a large scale or 'functional genomics' technique. They showed for the first time that mass spectrometry could in principle handle large numbers of proteins, similarly to oligonucleotide based methods. Shortly afterwards, two methods were introduced which allowed the use of peptide fragmentation data for much more specific identification of proteins [9–11].

During the last decade mass spectrometric instrumentation and procedures have been improved to a remarkable degree, increasing sensitivity of analysis and accuracy of results by orders of magnitude. This development seems set to continue in the future and will fuel additional advances.

In 1994, the term 'proteomics' was coined to designate the large scale characterization of the entire protein complement of a cell line, tissue or organism [12–14]. At the time, proteome analysis was mainly associated with the display of crude protein mixtures, such as tissue homogenates, on two-dimensional gels. Differences between the displayed spots in the normal and diseased state, for example, would be measured by image analysis of the stained protein spots and would correlate with the disease. A crucial and previously difficult step was the identification of such spots. In 1996 a first large scale protein identification project was performed which unambiguously demonstrated that mass spectrometry had the sensitivity, specificity and throughput to perform this task [15].

Recently, attention has shifted to 'functional proteomics', in which proteins are purified with a specific function in mind. Examples of such approaches are affinity purification to obtain binding partners of a protein of interest, purification of an organelle, or purification of a multi-protein complex with a defined function. This type of proteomics has become very successful and will continue to deliver answers to biological questions which would be very hard to obtain otherwise. Below, we will describe principles involved in the mass spectrometric analysis of proteins, review recent advances in the instrumentation and novel approaches in proteomics by mass spectrometry. Finally, we will illustrate some of the functional proteomic approaches with examples from our laboratory.

2. Mass spectrometry in proteomics and in functional genomics

In proteomics the premier task of mass spectrometry is the identification of very low levels of protein which have been

*Corresponding author. <http://www.pil.sdu.dk>.
E-mail: mann@bmb.sdu.dk

separated by one- or two-dimensional gel electrophoresis. Protocols for high sensitivity preparation of gel separated proteins have been published [16–18]. Proteins are first degraded into peptides by sequence specific proteases such as trypsin. Peptides are preferred for MS analysis, since proteins cannot easily be eluted from gels without detergents like SDS (which are detrimental to mass spectrometry) and because large proteins are usually heterogeneous and hence possess no single molecular weight which can be related to the corresponding entry in a sequence database. The masses of a set of peptides can be measured by matrix assisted laser desorption/ionization (MALDI) where a co-precipitate of light absorbing matrix (usually α -cyano-4-hydroxycinnamic acid or dihydroxy benzoic acid) and the peptide solution is irradiated with a short pulse of UV light in the vacuum. Some of the released peptides are ionized by attachment of protons and are accelerated in a strong electric field. After being turned around by an energy correcting ion mirror they are then detected on a channeltron detector. The result of this measurement is a time of flight distribution of the peptides in the supernatant of the trypsin digested protein (TOF mass spectrometry). After determining the peak centroids and calibrating the spectrum, for example on trypsin auto-digestion products, a set of highly accurate peptide masses is obtained. With state of the art MALDI mass spectrometers, peak resolution of about 10000 and a mass accuracy of a few parts per million is now possible. Recording the mass spectrum can be done manually in a few minutes and has also been entirely automated using fuzzy logic feedback control [19]. Fig. 1a shows the peptide mass map obtained fully automatically from a gel separated protein from the malaria parasite.

The MALDI peptide mass mapping, or ‘mass fingerprinting’ method has the advantage of being scalable. Many samples can be deposited on a single probe holder and measured in a single run. Together with automated and high sensitivity preparation of the proteins for analysis, hundreds of proteins can be handled in this format.

Unfortunately, not all proteins are amenable to identification by peptide mass mapping alone. A large percentage of human proteins are still not represented full length in sequence databases, small proteins sometimes do not result in a sufficient number of tryptic peptides for unambiguous identification and mixtures of proteins can only be ‘deconvoluted’ to their respective entries in the databases with special interpretation [20,21]. In many of those cases, a further step – mass spectrometric sequencing of the peptides – is required.

Electrospray mass spectrometry is a different but complementary method to identify proteins [1]. A solution containing the peptides is pumped through a metal capillary and dispersed at high voltage, resulting in rapidly evaporating, peptide containing droplets. After desolvation, the peptide molecules remain charged by attachment of one or a few protons and are drawn into the vacuum of a mass spectrometer. In the mass spectrometer, the ions are separated by dynamic electric fields according to their mass to charge ratios. After a mass spectrum is obtained, the instrument can be instructed to pass only a particular ion species into a ‘collision chamber’ where this ion species collides with nitrogen or argon gas at low pressure. Multiple impacts result in fragmentation of the peptide species, usually along the peptide bond. The C-terminal containing fragments are designated Y'' -ions and the N-terminal containing fragments are designated B-ions [22]. A series

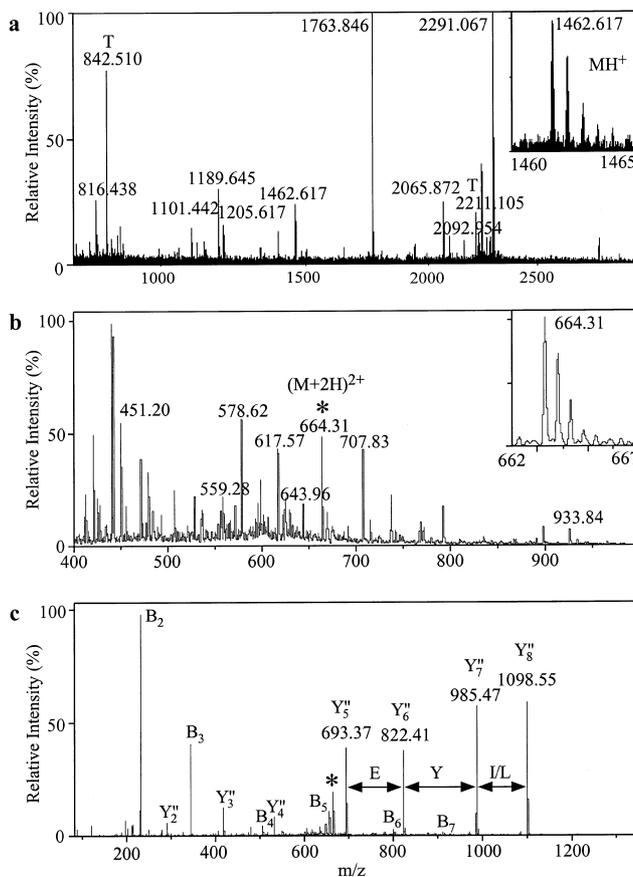


Fig. 1. a: Peptide mass map of a *Plasmodium falciparum* malaria protein obtained from an in-gel digest of a spot from a 2D gel. The list of masses unambiguously identified gene 14-3-3 in the database with 13 peptides matching the calculated tryptic peptide masses within 30 ppm. Two tryptic peptides (T) were used for calibration of the spectrum. (The proteins were analyzed using the automated analysis at Protana A/S, Odense, Denmark.) b and c: Identification of a human protein by nanoelectrospray mass spectrometry. b: Mass spectrum of the tryptic peptides from an in-gel digest. Marked peaks were fragmented and correspond to peptides from Nop5. c: Fragment mass spectrum of the doubly charged peptide ion at m/z 664.31. The sequence tag E-Y-I/L is derived from the mass difference of peptide fragments. The sequence, together with the start and end mass of the fragmentation series and the peptide mass, was combined into the search string (693.34)EYL(1098.52) and searched in a non-redundant database. The retrieved sequence TQLYEYLQNR was matched against the tandem mass spectrum. The position of the assigned series of N-terminal (B-ions) and C-terminal (Y'' -ions) are marked.

of B- or Y'' -ions spells out a partial sequence given by the molecular weight differences between the fragments (for example, a mass difference between three adjacent Y'' fragment ions at masses 500, 613, and 684 would spell out the sequence IG in the C- to N-terminal direction, where I could be isoleucine or leucine which have the same mass and cannot be distinguished).

In the last few years the triple quadrupole mass spectrometer, in which the first, isolating, mass spectrometer and the part of the mass spectrometer which separates the fragments were both of the quadrupole type, has largely been displaced by the quadrupole time of flight instrument. In this instrument, the fragments are detected by TOF mass spectrometry, like they are in the MALDI method. Resolution and mass

accuracy are likewise in the range of 10 000 and low ppm, respectively. Such high performance usually leads to unambiguous and straightforward identification on the basis of one or a few peptides which have been fragmented. Fig. 1b shows the mass spectrum of the peptides of a human protein. The peptide at mass to charge ratio of 664.31 was selected and fragmented, resulting in the fragment mass spectrum shown in Fig. 1c. The mass difference between the marked peaks spells the sequence EYI which, together with the mass of the peptide and the mass of the fragments, uniquely identified the sequence TQLYEYLQNR in the non-redundant database containing more than 350 000 proteins. Calculated fragments for that sequence fit 15 fragment peaks in the spectrum within 0.05 Da. Fragmentation of several additional peptides likewise uniquely identified the same protein in the database.

Peptide delivery from the gel supernatant after digestion of the protein can happen in two ways: In the ‘nanoelectrospray’ approach [23–26] the peptide mixture is micropurified in a small capillary packed with a small volume (100 nl) of reverse phased material. Peptides are eluted into a very fine tipped ‘spraying needle’ which supports sequencing runs of more than 30 min on 1 μ l of peptide solution. Peptides can be selected in turn for sequencing and peptide sequence tags can be obtained on a large number of peptides. Proteins can be identified in ‘real time’ during the experiment and the remainder of the sequencing time can be focused on minor components of a mixture or on potentially modified peptides. This approach has proven very robust and has helped solve a large number of important biological problems (see for example [27–30]). In the liquid chromatography/mass spectrometry approach (LC/MS), peptides are separated by a reversed phase chromatograph whose effluent is directly coupled to the mass spectrometer. Peptides are sequenced ‘on-line’ as they elute from the column [31–33]. Advantages of this approach include the potential to automate sample introduction and the fact that the peptides elute in a small volume, increasing the mass spectrometric response. LC/MS is often performed with an ‘ion trap’, a small relatively simple mass spectrometer which can automatically sequence and obtain tandem mass spectra at relatively high sensitivity [34]. Database searching is then often performed with correlating uninterpreted tandem mass spectra against the calculated mass spectra of all the peptides in the database [11,35].

Most mass spectrometric facilities can now identify proteins at about the 1 pmol level (50 to 100 ng of protein) of protein initially applied to the gel. Specialized laboratories have achieved sensitivities in the low ng range (weak silver stained levels). Mass spectrometry itself is exquisitely sensitive, requiring only a million or so molecules for detection. Therefore we believe that the impressive gain in sensitivity achieved over the last decade (more than a thousand-fold) will continue in the future as various bottlenecks in the identification process are addressed.

3. Novel developments in mass spectrometry and mass spectrometric approaches to proteomics problems

3.1. MALDI quadrupole time of flight instrument

An interesting instrument was developed recently, which combines advantages of the MALDI peptide mapping approach and the peptide sequencing method [36,37]. Briefly, a MALDI ion source is placed in front of a quadrupole TOF

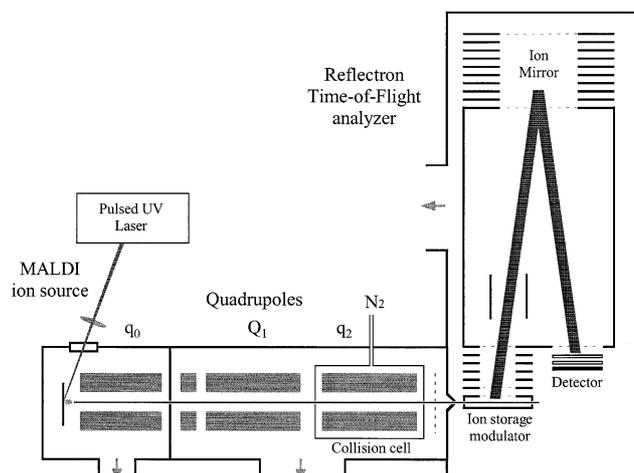


Fig. 2. Schematic of the MALDI quadrupole time of flight instrument. Ions are created by the MALDI process and are accelerated into the quadrupole sections. To obtain the peptide mass spectrum sections Q_0 , Q_1 and Q_2 are all operated in transmission mode and the ions are separated in the time of flight mass spectrometer. After obtaining the mass spectrum, an individual peptide species can be selected by section Q_1 . This ion species is fragmented by collision with argon or nitrogen gas in Q_2 . The fragments are analyzed with very high resolution and mass accuracy in the time of flight section. In a single experiment many peptide ion species can be sequenced.

instrument (MALDI QSTAR). Peptide mass maps are obtained by first passing the ions through the quadrupole section and then analyzing them by the TOF part. Individual peptide species can then be isolated, fragmented in the collision cell and analyzed by TOF. Fig. 2 shows a schematic of the instrument. In our experience, the MALDI QSTAR can achieve sub-picomole sensitivities on gel separated proteins, and identifications are very specific. Throughput is currently limited by the number of ions generated and transmitted. Developments now under way should increase this number at least ten-fold. We anticipate that the MALDI quadrupole time of flight instrument will play a major role in proteomics as the ‘work horse’ for protein identification.

3.2. Genome searching

To date, only protein sequence databases (usually auto-translated from DNA sequence data) and expressed sequence tag databases have been searched by mass spectrometric data. It has been shown that virtually all human proteins which can be visualized on gels can be identified in the expressed sequence tag databases, which now contain more than two million single read cDNA sequences [38]. However, ESTs usually cover only part of a protein sequence. If it were possible to work directly with genomic sequence databases, this would in principle allow for the identification of every peptide on which mass spectrometric sequence information was obtained. Difficulties in genome searching include the large size of the human genome and the fact that only a few percent are protein coding. Additionally, the exon–intron structure of most genes cannot be accurately predicted by bioinformatics [39,40]. Recent results obtained in our group [41], however, show that a suitably modified peptide sequence tag algorithm has enough specificity to uniquely locate peptides in raw genomic data. Furthermore, the mass spectrometric data helps in predicting the gene structure.

3.3. Analyzing crude mixtures of proteins without gel electrophoresis

Several interesting approaches have been taken recently towards the analysis of the proteome without the use of gel electrophoresis. In one such approach, the protein population is separated by a variant of capillary electrophoresis and the intact proteins are then eluted into a Fourier transform ion cyclotron resonance mass spectrometer (FT ICR). The FT ICR is capable of storing the ions and measuring them at extremely high resolution and mass accuracy using a frequency based method. Measurement of several hundreds protein components from lysates of *Escherichia coli* or yeast has already been shown [42]. Using a variant of the tandem mass spectrometric method, it may also be possible to identify the proteins 'on-line' as they elute into the mass spectrometer [43,44]. Several practical problems remain to be solved for this type of proteomic analysis to be successful. Furthermore, as mentioned above, larger proteins, even when they can be measured by this method, will have a molecular weight distribution rather than a single molecular weight.

In another approach, crude protein mixtures are digested in solution without separation. The resulting peptide mixture is then analyzed by the LC/MS method outlined above [45,46]. As the capacity of the mass spectrometer to sequence co-eluting peptides increases, more and more complex protein mixtures can be analyzed. It is clear, however, that in any direct analysis of very crude protein mixtures the major proteins will tend to mask the minor components. This 'dynamic range problem' is a key difficulty in any kind of unbiased analysis of crude mixtures such as tissue homogenates or cell lysates.

3.4. Quantitation by mass spectrometry

The signal intensity of the peptides in the mass spectrometer cannot directly be used to derive the quantity of the protein. Instead the staining intensity together with a rough estimate of the mass spectrometric response have usually been harnessed to determine the protein amount. Recently, isotopic labeling methods have been introduced into mass spectrometric proteome studies [47,48]. A stable isotope label is introduced into one of the two samples that are to be quantitated relatively to each other. These samples are then mixed and the ratio between peptide with isotopic label and without, accurately determines the ratio. Labels can for example be introduced through N15 media, through the blocking group employed in blocking reactive cysteines in the proteins or by derivatization of the N-termini of peptides. These techniques now open up for accurate quantitation in a wide range of proteomic situations.

4. Applications of mass spectrometry in proteomics

Mass spectrometry is now almost universally used as the identification method in 'expression proteomics' with two-dimensional gels of two different biological states. Such applications are not reviewed here but can be found, for example, in special issues of Electrophoresis as well as many recent reviews [49–55]. A common difficulty in such projects has been that it was difficult to extract biological mechanism from the up or down regulation of the abundant cellular proteins which are usually displayed in such maps. Furthermore, this approach is now superseded in many circumstances by the powerful oligonucleotide chip approaches [56]. Instead we will

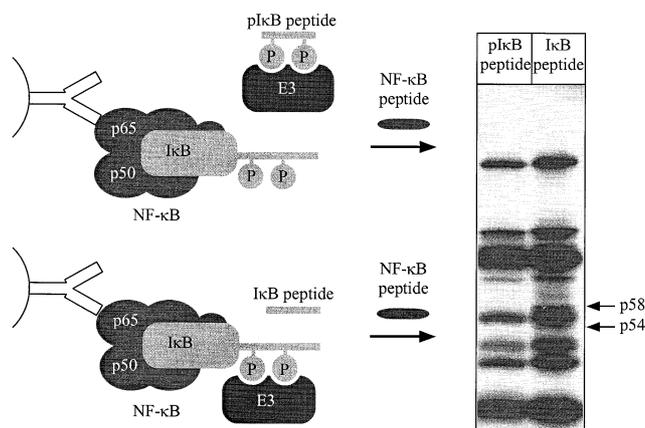


Fig. 3. Isolation of proteins associated with the phosphorylated IκB/NFκB protein complex. ^{35}S -radiolabeled HeLa lysate was incubated with antibody immobilized IκB/NFκB complex in the presence of phosphorylated IκB peptide (lane 1) or serine substituted IκB peptide (lane 2). The immune complexes were then washed, eluted and analyzed by SDS-PAGE and autoradiography. Tryptic peptides derived from the marked bands from both lanes were carefully investigated by nanoelectrospray tandem mass spectrometry and resulted in the identification of the E3 ligase receptor subunit of IκB [66].

focus on some successful examples from our laboratory of 'functional proteomics' in which organelle purification or affinity purification was used as a direct lead into biological function.

4.1. Multi-protein complexes

Multi-protein complexes, or 'molecular machines', are assemblies with a particular function such as splicing, transport or nuclear import/export. One use of proteomics technology is to determine the make up of such complexes [57,58]. To this end, they need to be purified specifically, the identity of the factors in the complex needs to be determined and finally the *in vivo* presence of the novel members of the complex needs to be established. In collaboration with Lamond's laboratory, we have extensively characterized the human spliceosome in this way. The spliceosome was assembled on a model biotinylated and radioactively labeled pre-messenger RNA. The complex was pulled out of the radioactive fraction after gel filtration using the biotin-streptavidin system. The spliceosome associated proteins were then separated and visualized by two-dimensional gel electrophoresis. More than 70 protein spots were analyzed. In a single experiment, 19 novel splicing factors were found. Interestingly, all the novel proteins could be identified, and many cloned in a straightforward manner, using EST libraries and physical EST clones as reagents [38]. *In vivo* confirmation of the involvement of these factors in splicing was obtained by fusing novel genes with green fluorescence protein (GFP) and observing co-localization with known spliceosomal protein. Several of the novel proteins are now further being studied in our laboratories with the help of proteomic techniques. For example, proteins can be epitope tagged and incubated with nuclear extract. The eluted and gel separated proteins can again be analyzed by mass spectrometry to reveal more detailed involvement in sub-complexes of the spliceosome, in this case.

A large number of multi-protein complexes has now been analyzed, including the nuclear pore complex in yeast [59], the spindle pole body in yeast [60], the Arp2/3 complex [61] and

many others. The yeast ribosome [62] and the ‘interchromatin granule’ [63] have been studied by digesting the full complement of proteins in solution and identifying as many proteins as possible by the LC/MS method mentioned above. Crucial issues in the large scale analysis of multi-protein complexes by mass spectrometry are the efficient tagging and purification of the ‘bait’ with the associated ‘prey’ proteins. Advances are being made in this area as well, for example by using two functionalities in the tag (calmodulin binding and Z-tag [64] or by using two epitope tags).

Chemical cross-linking is another method that can be employed to study multi-protein complexes. For example, our laboratory has performed limited cross-linking on the Nup85 sub-complex of the nuclear pore [65]. In addition to the proteins in the sub-complex additional bands appear which contain cross-linked proteins. Pairs or triplets of such proteins yielded nearest neighbor relationships which were used to build a crude topological model of the complex.

4.2. Signaling pathways

Proteomic methods have also been used in our laboratory to study more transient rather than structural complexes. Many signaling cascades are transmitted through multi-protein complexes involving scaffolds and these complexes can be biochemically purified. In collaboration with Ben-Neriah we have assembled the NF κ B signaling complex on beads to which NF κ B antibody was immobilized. When the NF κ B pathway is stimulated, the I κ Bs which normally protect the nuclear localization sequence in NF κ B are phosphorylated and then degraded by the ubiquitination machinery. The specific receptor which recognizes the phosphorylated epitope was not known at the time of this study. In stimulated, proteasome inhibited cells, the receptor was specifically visualized when the complex was incubated with a phosphorylated versus a non-phosphorylated synthetic peptide corresponding to the sequence of I κ B which is phosphorylated upon stimulation (see Fig. 3). One-dimensional gels revealed subtle differences in eluted proteins and the corresponding bands were sequenced by nano-electrospray mass spectrometry. Six different

proteins were identified in the band and the control but two of the peptides were absent in the control. These identified a *Drosophila* protein and a human EST, leading to the cloning of the receptor (now named E3RS^{I κ B}) [66].

In a different approach, we have precipitated tyrosine phosphorylated proteins in the epidermal growth factor receptor (EGFR) pathway. Cells were stimulated by EGF and phosphorylated proteins were precipitated by anti-phosphotyrosine antibodies [67]. Nine proteins specific to this pathway were identified by comparison of the immuno-precipitates between stimulated and non-stimulated cells. Interestingly, even in such a well studied pathway, one protein, vav-2, was found which was not previously known to be involved and a completely novel protein was cloned. Note that in this procedure, proteins are precipitated both from the complex that forms directly at the intracellular domain of the receptor as well as members of the cascade that are further downstream and do not have physical contact to the receptor.

4.3. Organelles

Apart from multi-protein complexes, organelles can also be purified and their composition analyzed by mass spectrometry. Since organelles are often less well defined than smaller multi-protein complexes, the task of verification of identifications becomes even more important.

Working with the laboratory of Lamond we are purifying and identifying the components of various nuclear structures. Fig. 4 shows a one-dimensional gel of human nucleolar proteins purified from HeLa cells. Slicing the gel into pieces and analyzing the main protein components in every slice by MALDI mass fingerprinting and nano-electrospray peptide sequencing identified more than one hundred proteins. Interestingly, many of the proteins did not appear on two-dimensional gel of the same preparation. Some of the proteins were too large or too hydrophobic for two-dimensional gel analysis but others also appear to have been too basic for the range of isoelectric focusing used in our gel system (pI 3–10). We anticipate that many organelles will now be purified and analyzed by mass spectrometry. For the first time it will be pos-

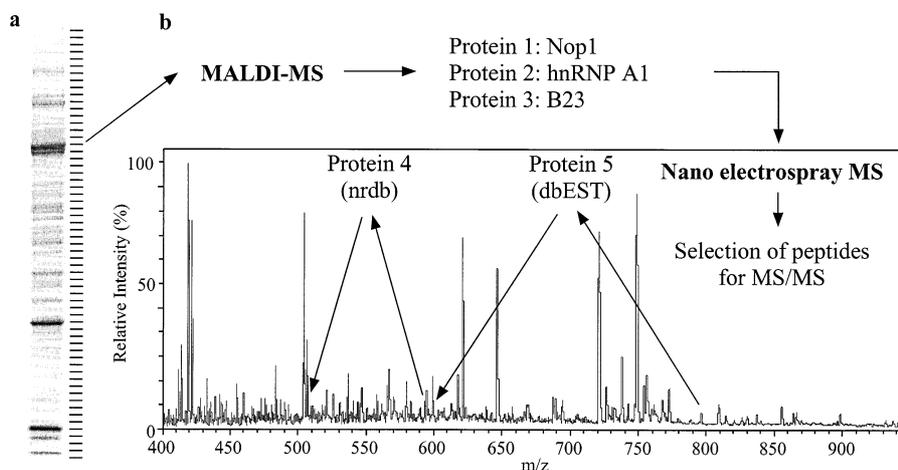


Fig. 4. Characterization of proteins from human nucleoli. a: 1D gel of nucleolar proteins. MALDI MS of an aliquot of the tryptic peptide mixture from one of the slices resulted in the identification of Nop1, hnRNP A1 and B23. b: Nano-electrospray mass spectrum of the same peptide mixture. The major signals represent peptides from the three proteins already identified by MALDI MS. Sequencing of unexplained peptides resulted in the identification of two additional proteins. The sequences were retrieved from different databases in real time during the nano-electrospray experiment. Sequencing of additional peptides predicted from the retrieved sequences therefore can be used to validate the identification. Nano-electrospray of peptide mixtures has a unique advantage for such an approach as compared to LC/MS because any peptide can be selected for sequencing at any time during the experiment.

sible to obtain an inventory of many of the basic structures of the cell. Even more intriguingly, after the basic catalogue of protein components is established, it is then possible to perturb the system to directly measure protein changes such as recruitment of novel factors or bridging to other complexes and structures.

5. Conclusion

As we have shown in this review, mass spectrometry is the central technique in a wide variety of functional genomics, or proteomics approaches to study gene function in the post-genomics world. Mass spectrometric instrumentation continues to become more powerful and novel instrumental concepts are being put into use. The imminent completion of the human genome will allow all human proteins to be correlated directly with their corresponding database entries. A wide variety of approaches to study protein function now use mass spectrometry. Most of these are affinity based, which simultaneously addresses the problem of functional significance and the dynamic range problem in proteomics. Mass spectrometry is sure to have a place as one of the fundamental techniques in functional genomics in the developing era of systematic biology.

Acknowledgements: All members of the Protein Interaction Laboratory are gratefully acknowledged for valuable discussions and comments on the manuscript. We are also thankful to our collaborators mentioned in this review, particularly Angus Lamond, Dundee and Yinon Ben-Neriah, Jerusalem, and Protana A/S. This work was funded by a generous grant of the Danish National Research Foundation to Matthias Mann's laboratory at the Center for Experimental Bioinformatics (CEBI).

References

- [1] Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. (1989) *Science* 246, 64–71.
- [2] Karas, M. and Hillenkamp, F. (1988) *Anal. Chem.* 60, 2299–2301.
- [3] Hillenkamp, F., Karas, M., Beavis, R.C. and Chait, B.T. (1991) *Anal. Chem.* 63, 1193A–1202A.
- [4] Henzel, W.J., Billeci, T.M., Stults, J.T. and Wong, S.C. (1993) *Proc. Natl. Acad. Sci. USA* 90, 5011–5015.
- [5] Mann, M., Højrup, P. and Roepstorff, P. (1993) *Biol. Mass Spectrom.* 22, 338–345.
- [6] Pappin, D.J.C., Højrup, P. and Bleasby, A.J. (1993) *Curr. Biol.* 3, 327–332.
- [7] James, P., Quadroni, M., Carafoli, E. and Gonnet, G. (1993) *Biophys. Biochem. Res. Commun.* 195, 58–64.
- [8] Yates, J.R., Speicher, S., Griffin, P.R. and Hunkapiller, T. (1993) *Anal. Biochem.* 214, 397–408.
- [9] Mann, M. (1993) presented at the 7th Symposium of the Protein Society, San Diego.
- [10] Mann, M. and Wilm, M.S. (1994) *Anal. Chem.* 66, 4390–4399.
- [11] Eng, J.K., McCormack, A.L. and Yates, III, J.R. (1994) *J. Am. Soc. Mass Spectrom.* 5, 976–989.
- [12] Wasinger, V.C. et al. (1995) *Electrophoresis* 16, 1090–1094.
- [13] Wilkins, M.R. et al. (1996) *BioTechnology* 14, 61–65.
- [14] Wilkins, M.R., Williams, K.L., Apple, R.D. and Hochstrasser, D.F. (1997) *Proteome Research: New Frontiers in Functional Genomics*, p. 243, Springer, Berlin.
- [15] Shevchenko, A. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93, 14440–14445.
- [16] Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996) *Anal. Chem.* 68, 850–858.
- [17] Jensen, O., Wilm, M., Shevchenko, A. and Mann, M. (1999) *Methods Mol. Biol.* 112, 513–530.
- [18] Pandey, A., Andersen, J. and Mann, M. (2000) *STKE*, Science online.
- [19] Jensen, O.N., Mortensen, P., Vorm, O. and Mann, M. (1997) *Anal. Chem.* 69, 1706–1714.
- [20] Jensen, O.N., Podtelejnikov, A.V. and Mann, M. (1997) *Anal. Chem.* 69, 4741–4750.
- [21] Fenyo, D., Qin, J. and Chait, B. (1998) *Electrophoresis* 19, 998–1005.
- [22] Roepstorff, P. and Fohlman, J. (1984) *Biomed. Mass Spectrom.* 11, 601.
- [23] Wilm, M.S. and Mann, M. (1994) *Int. J. Mass Spectrom. Ion Process.* 136, 167–180.
- [24] Wilm, M. and Mann, M. (1996) *Anal. Chem.* 68, 1–8.
- [25] Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T. and Mann, M. (1996) *Nature* 379, 466–469.
- [26] Jensen, O.N., Wilm, M., Shevchenko, A. and Mann, M. (1998) in: *2-D Proteome Analysis Protocols* (Link, A., Ed.), Vol. 112, pp. 571–588, Humana Press, Totowa.
- [27] Muzio, M. et al. (1996) *Cell* 85, 817–827.
- [28] Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. and Cech, T.R. (1997) *Science* 276, 561–567.
- [29] Mercurio, F. et al. (1997) *Science* 278, 860–866.
- [30] Ciosk, R., Zachariae, W., Michaelis, C., Shevchenko, A., Mann, M. and Nasmyth, K. (1998) *Cell* 93, 1067–1076.
- [31] Moseley III, M.A., Deterding, L.J., Tomer, K.B. and Jorgenson, J.W. (1991) *Anal. Chem.* 63, 1467–1473.
- [32] Hunt, D.F. et al. (1992) *Science* 255, 1261–1263.
- [33] Haynes, P.A., Fripp, N. and Aebersold, R. (1998) *Electrophoresis* 19, 939–945.
- [34] Jonscher, K.R. and Yates, J.R. (1997) *Anal. Biochem.* 244, 1–15.
- [35] Yates 3rd, J.R., Carmack, E., Hays, L., Link, A.J. and Eng, J.K. (1999) *Methods Mol. Biol.* 112, 553–569.
- [36] Krutchinsky, A.N., Zhany, W. and Chait, B.T. (2000) *J. Am. Soc. Mass Spectrom.* 6, 493–504.
- [37] Shevchenko, A., Loboda, A., Shevchenko, A., Ens, W. and Standing, K.G. (2000) *Anal. Chem.* 72, 2132–2141.
- [38] Neubauer, G. et al. (1998) *Nat. Genet.* 20, 46–50.
- [39] Krogh, A. (1998) in: *Guide to Human Genome Computing* (Bishop, M.J., Ed.), pp. 261–274, Academic Press, San Diego.
- [40] Dunham, I. et al. (1999) *Nature* 402, 489–495.
- [41] Küster, B., Mortensen, P. and Mann, M. (1999) in: *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, pp. 1897–1898, Dallas, TX.
- [42] Jensen, P.K., Pasa-Tolic, L., Anderson, G.A., Horner, J.A., Lipton, M.S., Bruce, J.E. and Smith, R.D. (1999) *Anal. Chem.* 71, 2076–2084.
- [43] Mørtz, E., O'Connor, P.B., Roepstorff, P., Kelleher, N.L., Wood, T.D., McLafferty, F.W. and Mann, M. (1996) *Proc. Natl. Acad. Sci. USA* 93, 8264–8267.
- [44] Li, W., Hendrickson, C.L., Emmett, M.R. and Marshall, A.G. (1999) *Anal. Chem.* 71, 4397–4402.
- [45] Yates, J., McCormack, A., Schieltz, D., Carmack, E. and Link, A. (1997) *J. Protein Chem.* 16, 495–497.
- [46] Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M. and Yates, J.R. (1999) *Nat. Biotechnol.* 17, 676–682.
- [47] Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R. (1999) *Nat. Biotechnol.* 17, 994–999.
- [48] Oda, Y., Huang, K., Cross, F.R., Cowburn, D. and Chait, B.T. (1999) *Proc. Natl. Acad. Sci. USA* 96, 6591–6596.
- [49] Celis, J. et al. (1996) *FEBS Lett.* 398, 129–134.
- [50] Celis, J.E. et al. (1999) *Electrophoresis* 20, 300–309.
- [51] Jungblut, P.R. et al. (1999) *Electrophoresis* 20, 2100–2110.
- [52] Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S. and Garrels, J.I. (1999) *Mol. Cell. Biol.* 19, 7357–7368.
- [53] Gauss, C., Kalkum, M., Lowe, M., Lehrach, H. and Klose, J. (1999) *Electrophoresis* 20, 575–600.
- [54] Pennington, S.R. et al. (1997) *Trends Cell Biol.* 7, 168–173.
- [55] Page, M.J. et al. (1999) *Proc. Natl. Acad. Sci. USA* 96, 12589–12594.
- [56] Celis, J.E. et al. (2000) *FEBS Lett.* 480, 2–16.
- [57] Lamond, A.I. and Mann, M. (1997) *Trends Cell Biol.* 7, 139–142.
- [58] Neubauer, G., Gottschalk, A., Fabrizio, P., Séraphin, B., Lührmann, R. and Mann, M. (1997) *Proc. Natl. Acad. Sci. USA* 94, 385–390.

- [59] Rout, M.P., Aitchison, J.D., Suprapto, A., Hjertaas, K., Zhao, Y. and Chait, B.T. (2000) *J. Cell Biol.* 148, 635–651.
- [60] Wigge, P.A., Jensen, O.N., Holmes, S., Soues, S., Mann, M. and Kilmartin, J.V. (1998) *J. Cell Biol.* 141, 967–977.
- [61] Winter, D., Podtelejnikov, A.V., Mann, M. and Li, R. (1997) *Curr. Biol.* 7, 519–529.
- [62] Tong, W., Link, A., Eng, J.K. and Yates, J.R. (1999) *Anal. Chem.* 71, 2270–2278.
- [63] Mintz, P.J., Patterson, S.D., Neuwald, A.F., Spahr, C.S. and Spector, D.L. (1999) *EMBO J.* 18, 4308–4320.
- [64] Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Seraphin, B. (1999) *Nat. Biotechnol.* 17, 1030–1032.
- [65] Rappsilber, J., Siniosoglou, S., Hurt, E.C. and Mann, M. (2000) *Anal. Chem.* 72, 267–275.
- [66] Yaron, A. et al. (1998) *Nature* 396, 590–594.
- [67] Pandey, A., Podtelejnikov, A.V., Blagoev, B., Bustelo, X.R., Mann, M. and Lodish, H.F. (2000) *Proc. Natl. Acad. Sci. USA* 97, 179–184.