Ann Ist Super Sanità 2005;41(4):479-482

Homo sapiens proteomics: clinical perspectives

Piero DEL BOCCIO and Andrea URBANI

Centro Studi sull'Invecchiamento, Dipartimento di Scienze Biomediche, Università degli Studi "G. D'Annunzio" di Chieti e Pescara, Chieti, Italy

Summary. - The complete human genome sequencing project opened the route to a new holistic scientific movement, characterised by the investigation of the functional elements of our genetic heredity mainly in respect to multifactorial diseases such as cancer, diabetes, dementias, etc. The clinical proteomics tries to highlight the relationships between the main molecular physiological effectors, proteins, with defined physiopathological conditions. The development of a new generation of mass spectrometers is deeply changing the technologies employed in these investigations, enabling direct quantitative analysis of clinical samples such as, blood, urine, saliva and biopsy samples. In this manuscript we introduce and discuss the main phases of these investigations and their possible follow up in diagnostic and functional studies.

Key words: molecular biomarker, MALDI-TOF-MS, mass spectrometry, proteomics.

Riassunto (*Proteomica dell'*homo sapiens*: prospettive cliniche*). - A seguito del sequenziamento del genoma umano un nuovo movimento scientifico olistico ha caratterizzato lo studio degli elementi funzionali della nostra eredità genetica, in particolare nella sua relazione verso malattie multifattoriali quali, diabete, cancro, demenze, ecc. La proteomica clinica cerca di mettere in luce le correlazioni tra i principali effettori molecolari delle nostre funzioni vitali, le proteine, con definiti quadri fisiopatologici. L'avvento di spettrometri di massa di nuova concezione sta trasformando profondamente questo settore aprendo le porte ad analisi quantitative direttamente su campioni clinici quali: sangue, urine, saliva, campioni bioptici. In questo manoscritto cercheremo di presentare e discutere le fasi salienti di queste indagini e le loro ricadute sia diagnostiche che funzionali.

Parole chiave: marcatori molecolari, MALDI-TOF-MS, spettrometria di massa, proteomica.

Introduction

Following the great promise of the human genome project, a plethora of novel studies have been carried out to investigate the molecular mechanisms of the encoded genes. The main interest of clinical research has consequently shifted toward an improved understanding of the inter-relations of functional elements and transcribed genome regions with specific physiopathologic conditions. This enthusiasm brought novel technological breakthroughs as much as intellectual developments in the experimental design. These were particularly driven by an holistic view of biological phenomena, which are now considered the consequence of organic wholes that are greater than the simple sum of their components. In fact, the human genome project provided us with the first holistic view of our genetic repository, shading new light in the inductive investigation of other molecular and functional repertories such as transcripts, proteins, metabolites, molecular interactions, etc.

Proteome investigations, called proteomics, are the natural development of this approach, now encompassing not only all the proteins encoded in any given cell, but also the set of all proteins isoforms and modifications, the structural description of proteins and their higher-order complex. As proteins are the main effectors of physiological functions, proteomics spans over a wide area of the present "post genomic" investigations. This high-throughput biochemistry will be crucial for the emerging field of system biology, contributing at a direct level to a full description of cellular function and biological phenomena [1, 2]. However, the sequencing of the human genome and that of numerous pathogens has opened the door for proteomics by providing a sequence-based framework for mining proteomes. Most difficult problems ahead is to find out how genes contribute to diseases that have complex pattern of inheritance, such as in the cases of diabetes, asthma, cancer and mental illness. In all these cases, no single gene has implicated in the disease process obtaining a dramatic complex situation.

Indirizzo per la corrispondenza (Address for correspondence): Andrea Urbani, Centro Studi sull'Invecchiamento, Università degli Studi di Chieti e Pescara, Via Colle dell'Ara, 66100 Chieti Scalo, Italy. E-mail: a.urbani@unich.it.

There is high interest in applying proteomics to foster a better understanding of disease dynamics. In particular clinical biochemistry has found a new era taking advantage from this holistic point of view. The interest in clinical proteomics is due in part to the prospects that a proteomics approach to disease investigations will overcome some of the limitations of other approaches. Because most drug targets are proteins, it is inescapable that proteomics will promote drugs discovery by the identification of new targets for therapeutics with more effective strategies for clinical practice [3].

Mass spectrometry

The detection of proteins profiles associated with disease states dates back to the very beginning of proteomics, when two-dimensional gel electrophoresis was first applied to clinical material. The advent of novel mass spectrometers enable to resolve thousands of protein and peptide species in body fluids is set to revolutionize protein-based diagnostic. In fact, mass spectrometry has increasingly become the method of choice for analysis of complex protein samples establishing itself as an indispensable technology to interpret the information encoded in genomes. This new era started with the discovery and development of protein ionization methods, as recognised by the 2002 Nobel prize in chemistry [4]. The ability of mass spectrometry to identify and accurately quantify even small amount of a specific protein from increasingly complex mixtures is becoming a primary driving force in proteomics. The rapid development of this technology has carried out instrument with an increased sensitivity, robustness and data handling with improved analytical performance.

Direct mass spectrometry approach of complex mixtures, is today, the last challenge for gel-free analysis of disease biomarkers. In particular, Linear Matrix-Assisted Laser Desorption-Ionization Time-of-Flight mass spectrometry (MALDI-TOF-MS), is a promising technique for clinical applications given its high sensitivity, accuracy and resolution in discriminating the low molecular weight protein profiles [5, 6]. Quantitative data are achievable by the use of internal standard gaining precision and accuracy within 20-25% . Moreover time and cost of analysis for each sample are particularly low, given the ion source set up, which eventually might handle hundreds of samples in a single acquisition batch. These qualities have made MALDI-TOF-MS as an instrument of choice to investigate large number of clinical relevant molecules in serum, blood, urine, tissue extracts and whole cells [7, 8]. Linear MALDI-TOF-MS coupled to surface activated sample targets (SELDI-TOF-MS) has already been recently employed by several groups to detect potential novel biomarkers in different biological samples [9-11]. Approaches include comparative analysis of protein expression in normal and disease samples to identify aberrantly expressed proteins that may represent new disease-markers. Serum patterns that distinguish between disease and normal subject with remarkable accuracy have been reported for several type of cancers and other diseases [12]. This direct mass spectrometry approach yield to a comprehensive profiles of peptides and proteins in biological fluids without the need to first carry out proteins separations and requiring reduced amount of sample for high throughput. Researchers are employed to develop new and automated peptide extractions, crystallizations, spectral acquisition and signal processing. Computational methodologies and bioinformatics tools are also necessary to develop models to analyze and compare large numbers of data points, essential for valuable and significant data mining (Fig. 1).

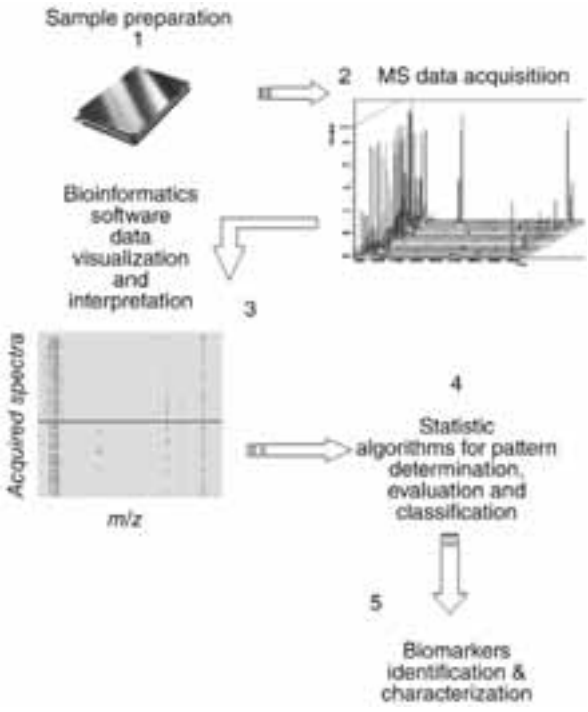

Fig. 1. - Experimental design of biological samples directly analyzed by MALDI-TOF-MS. 1) Pre-analytical step where the biological samples are processed and deposited on the MALDI target. 2) Complete screening of the samples derived from pathological and healthy subjects using linear MALDI-TOF-MS technology. 3-4) Combined visualization and statistical features for an easy and efficient identification of biomarker candidates. 5) Identification of the new specific disease-targets.

Experimental strategies

In Fig. 1 the complete procedure and data interpretation of biological samples directly analyzed by MALDI-TOF-MS was shown. The typical experiment consist of four stages. Stage 1 is a pre-analytical step where the biological sample are processed and deposited on the MALDI target. This often includes internal standards added to validate the analytical assessment and μSPE extraction (usually Reverse-Phases). In stage 2 the complete screening of the samples derived from pathological and healthy subjects using linear MALDI-TOF-MS technology are presented. In stages 3 and 4 the combined visualization and statistical features allows for an easy and efficient identification of biomarker candidates. Such candidates can later be used in pattern recognition models to analyze independent data to the significance of these tentative biomarkers in an open translational investigation for para-clinical studies. The last step of this process should be the identification of the outlined specific molecular disease marker. This investigation is necessary to pursued the potential functional insight of the outlined marker. Moreover, these data will provide a rational basis on their role in the pathological process, possibly suggesting new pharmacological protocols.

As shown in Fig. 2, the identification of the proteins corresponding to the disease-discriminating signals can be carry out by a RP-HPLC-MS separation to collect purified fractions containing the candidate biomarkers. Samples, generally, need to be preliminary pre-purified (ultrafiltration, SPE, etc.) obtaining a fraction in the interesting mass range. This liquid fraction can be directly injected and separated for fraction collection. Each purified liquid fraction are spotted on the MALDI target and analyzed in order to select only those with the interest signals. Finally the purified liquid fractions selected are digested with trypsin leading to peptides with C-terminally protonated amino acids, providing a direct advantage in subsequent mass spectrometry analyses. Protein identification can be carry out by MALDI-TOF Peptide Mass Fingerprint (PMF) analysis and by LC-ESI-MS/MS fragmentation analysis. The last approach consist of one or more step of high-performance liquid chromatography in micro-bore capillaries and elution into an electrospray ion source where the peptides are ionized and successively fragmented obtaining a series of MS/MS spectra. MS/MS approach is more efficient in this case considering the separation step less efficient than 2D-gel methodology usually coupled with PMF analysis. Proteins are identified by matching a list of experimental MS and MS/MS data with the calculated list of all data masses of each entry in a dadabase (for example a comprehensive protein database). By avoiding peak identification, the resulting classification may produce an algorithm suitable for prediction but ignoring the molecular basis of this classification. Such strategy could be dangerous since it might lead to a group discrimination based on a stochastic association of events rather then on the clinical groups investigated. Moreover, signals identification and characterization might return useful information on the aetiology and molecular mechanism of the investigated phenomena.

The identification of the candidate biomarkers is giving an important functional insight, possibly shading a light on the implications in the pathological process.

Future perspectives

An innovative mass spectrometry-diagnostic approach developed in the last years is called "MALDI-TOF imaging mass spectrometry" (MALDI-IMS) representing one of the most exciting and promising challenge in the field of diagnostic based-mass spectrometry [13]. This very recent technology exploits the methodology of MALDI-TOF MS combined with the direct employment of histological sections, therefore returning a direct molecular analysis from raw samples [14].

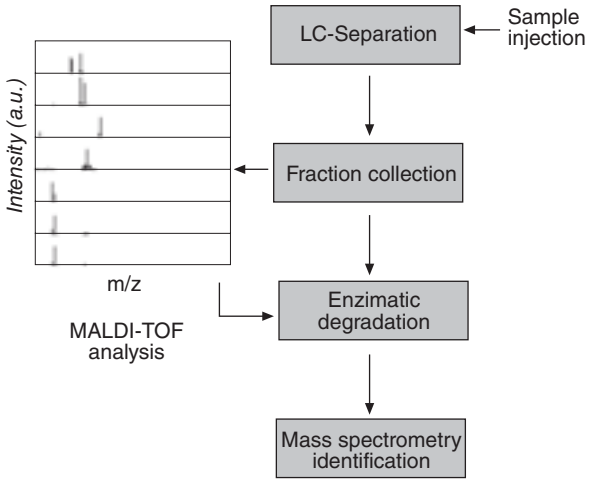

Fig. 2. - Complete procedure for the identification of the proteins corresponding to the disease-signal, by a RP-HPLC-MS separation to collect purified fractions containing the candidate biomarkers. Each purified liquid fraction is spotted on the MALDI target and analyzed in order to select only those with the interest signals. The selected purified liquid fractions, are digested generically with trypsin. Protein identification can be carried out by MALDI-TOF Peptide Mass Fingerprint (PMF) analysis and by LC-ESI-MS/MS fragmentation analysis.

MALDI-IMS provides profiles and two dimensional ion density maps of molecules directly from raw tissue sections, showing the relative abundance and spatial distribution of molecules. Consequently, it gives a protein profile expression with high sensitivity of an histological well defined tissue area, avoiding the loss of some important information coming up from the classic analytical separation methodologies (Fig. 3). Subsequently, proteins identification might be pursued using a classic approach as described for biological fluid investigations.

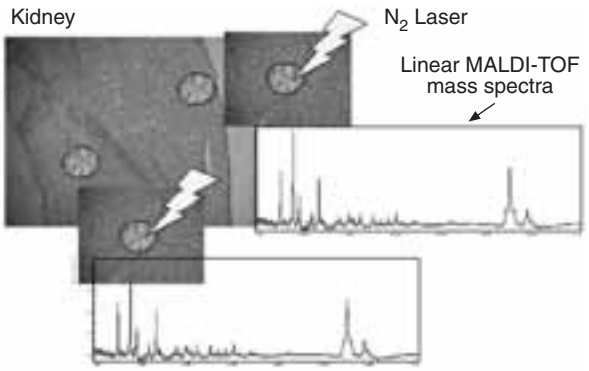

Fig. 3. - MALDI-IMS with the direct employment of histological sections of kidney, returning a two dimensional ion density maps of molecules directly from raw tissue sections.

In conclusion, MALDI-TOF-MS pattern-profiling will undoubtedly attain a prominent and lasting position in the future of the diagnostic. However, today the pattern-profiling proteomics methodology may not be quite reproducible, because several line of evidence indicate that there are a number of clinical and analytical chemistry factors that are major sources of variability and bias [15] (blood samples collection, serum preparation, storage and handling, sample extraction, etc.). Furthermore, human plasma and serum proteomics may be particularly susceptible to observational biases because many possible confounding factors such as smoking, diet or ascertainment bias. These could conceivably cause a phenotypic response that may be confused with a specific characteristic of the disease process under investigation. Future perspectives focus in standardizing samples preparation and preservation, reliable methods validation procedure in order to evaluate the significance of mass spectrometry-profiling proteomics in the clinical practices.

Receveid on 5 September 2005.
Accepted on 3 November 2005.

REFERENCES

1. Tyres M, Mann M. From genomics to proteomics. *Nature* 2003;422(6928):193-7.
2. Csete ME, Doyle JC. Reverse engineering of biological complexity. *Science* 2002;295(5560):1664-9.
3. Hanash S. Disease proteomics. *Nature* 2003;422(6928):226-32.
4. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422(6928):198-207.
5. Kallweit U, Börnsen OK, Kresbach GM, Widmer, HM. Matrix compatible buffers for analysis of proteins with matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 1996;10(7):845-9.
6. Dubois F, Knochenmuss R, Steenvoorden RJ, Breuker K, Zenobi R. On the mechanism and control of salt-induced resolution loss in matrix-assisted laser desorption/ionization. *Eur J Mass Spectrom* 1996;2(10):167-72.
7. Hutchens TW, Yip TT. New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom* 1993;7:576-80.
8. Veenstra TD, Conrads TP, Hood BL, Avellino AM, Ellenbogen RG, Morrison RS. Biomarkers: mining the biofluid proteome. *Mol Cell Proteomics* 2005;4(4):409-18.
9. Dekker LJ, Boogerd W, Stockhammer G, Dalebout JC, Siccama I, Zheng P, Bonfrer JM, Verschuuren JJ, Jenster G, Verbeek MM, Luider TM, Smith PA. MALDI-TOF mass spectrometry analysis of cerebrospinal fluid tryptic peptide profiles to diagnose leptomeningeal metastases in patients with breast cancer. *Mol Cell Proteomics* 2005;4(9):1341-9.
10. Carette O, Demalte I, Scherl A, Yalkinoglu O, Corthals G, Burkhard P, Hochstrasser DF, Sanchez JC. A panel of cerebrospinal fluid potential biomarkers for the diagnosis of Alzheimer's disease. *Proteomics* 2003;3(8):1486-94.
11. Sanchez JC, Guillaume E, Lescuyer P, Allard L, Carette O, Scherl A, Burgess J, Corthals GL, Burkhard PR, Hochstrasser DF. Cystatin C as a potential cerebrospinal fluid marker for the diagnosis of Creutzfeldt-Jakob disease. *Proteomics* 2004;4(8):2229-33.
12. Petricoin EF, Zoon KC, Kohn EC, Barret JC, Liotta LA. Clinical proteomics: translating benchside promise into bedside reality. *Nature Rev Drug Discov* 2002;1(9):683-95.
13. Maddalo G, Petrucci F, Iezzi M, Pannellini T, Del Boccio P, Ciavardelli D, Biroccio A, Forli F, Di Ilio C, Ballone E, Urbani A, Federici G. Analytical assessment of MALDI-TOF imaging mass spectrometry on thin histological samples. An insight in proteome investigation. *Clin Chim Acta* 2005;357(2):210-8.
14. Caldwell RL, Caprioli RM. Tissue profiling by mass spectrometry: a review of methodology and applications. *Mol Cell Proteomics* (2005);4:394-401. Review.
15. Villanueva J, Philip J, Chaparro CA, Li Y, Toledo-Crow R, DeNoyer L, Fleisher M, Robbins RJ, Tempst P. Correcting common errors in identifying cancer-specific serum peptide signatures. *Journal of Proteome Research* 2005;4(4):1060-72.