Lien du Congrès: http://www.biotech-ecolo.net

Proteomics technologies and approaches for the identification of disease biomarkers: Application to cancer and infectious diseases

HAOUDI Abdelali

Eastern Virginia Medical School, Department of Microbiology and Molecular Cell Biology, George L. Wright Center for Biomedical Proteomics, Lewis Hall #3011, Norfolk, VA 23501, USA, Email: Haoudia@evms.edu

Abstract

Proteomics is the study of the function of all expressed proteins. Tremendous progress has been made in the past few years in generating large-scale data sets for protein–protein interactions, organelle composition, protein activity patterns and protein profiles in cancer patients. Proteomics complements other functional genomics approaches, including microarray-based expression profiles, systematic phenotypic profiles at the cell and organism level, systematic genetics and small-molecule-based arrays. The ability of mass spectrometry to identify ever smaller amounts of protein from increasingly complex mixtures is a primary driving force in proteomics.

We have used mass-spectrometry based proteomic chip technology in order to define differential protein expression profiles in sera from HTLV-1 (Human T Cell Leukemia Virus Type 1)-infected individuals and determine the resulting proteomic changes that define ATL (HTLV-1 –associated Adult T Cell Leukemia) and HAM/TSP (HTLV-1 –associated neurodegenerative disease).

HTLV-1 is estimated to currently infect close to 20 million people worldwide. Our approach is to develop disease-specific protein expression profiles, which can be used as predictors of disease outcome and as biomarkers.

INTRODUCTION

The term proteome defines the entire protein complement in a given cell, tissue or organism. In its wider sense, proteomics research also assesses protein activities, modifications and localization, and interactions of proteins in complexes. It is very much a technology driven enterprise.

During the early years of proteomics and until relatively recently, profiling of protein expression in disease relied primarily on the use of two-dimensional polyacrylamide gel electrophoresis (2D PAGE), which was later combined with mass spectrometry. Most studies of this nature followed an approach in which a cocktail was used to solubilize the protein contents of an entire cell population, tissue or biological fluid, followed by separation of the protein contents of the lysate using 2D gels and visualization of the separated proteins using silver staining. It became clear that such an approach allows only a limited display of protein content that consisted of relatively abundant proteins. Nevertheless, profiling of disease tissues using this approach has had some utility. For example, it was demonstrated long before the use of DNA microarrays that leukaemias could be classified into their different subtypes using 2D PAGE.

Mass spectrometry (MS) has increasingly become the method of choice for analysis of complex protein samples (Aebersold and Mann, 2003). So far, protein analysis (primary sequence, post-translational modifications (PTMs) or protein–protein interactions) by MS has been most successful when applied to small sets of proteins isolated in specific functional contexts. The systematic analysis of the much larger number of proteins expressed in a cell, an explicit goal of proteomics, is now also rapidly advancing, due mainly to the development of new experimental approaches. Proteomics still faces significant technical challenges.

Each breakthrough that either allows a new type of measurement or improves the quality of data made by traditional types of measurements

expands the range of potential applications of MS to molecular and cellular biology.

By definition, a mass spectrometer consists of an ion source, a mass analyser that measures the mass-to-charge ratio (*m*/*z*) of the ionized analytes, and a detector that registers the number of ions at each *m*/*z* value. Electrospray ionization (ESI), matrix-assisted laser desorption/ionization (MALDI) and surface enhanced laser desorption/ionization (SELDI) (Figure 1) are the techniques most commonly used to volatize and ionize the proteins or peptides for mass spectrometric analysis (Fen et al., 1989; Wright, 2002). They can be combined with a variety of mass analyzers (TOF, quadrupole, ion trap and Fourier transform ion cyclotron).

Proteins are either identified by searching databases with the masses of proteolytic peptides (peptide mass fingerprinting) or using fragmentation data (raw MS/MS spectra or sequence tags).

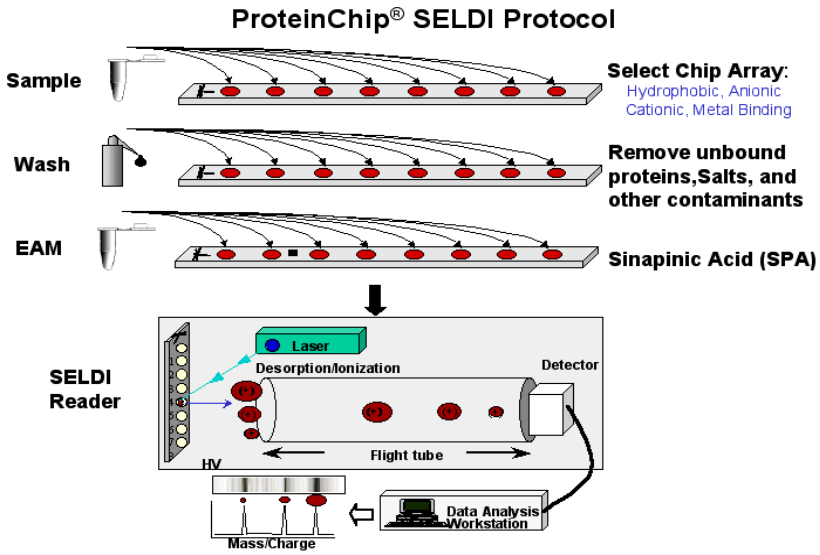

Figure 1. The ProteinChip SELDI-TOF-MS process. Sample proteins are applied to the chip using desired chemistry. The samples are coated desorption initiated by ionization and the time of flight recorded. This information is converted to protein mass. The process is highly automated and designed for high through put analysis.

Protein mixtures of considerable complexity can now be routinely characterized in some depth using the methods described above. One measure of technical progress is the number of proteins identified in each study. Such numbers can now reach into the thousands for suitably complex samples. But to be biologically useful, as opposed to simply highlighting analytical features of the methods, large-scale proteomic studies need to solve biological questions. In this regard, MS-based proteomics has interfaced particularly well with three types of biological or clinical questions. The first is the generation of protein–protein linkage maps. The second is the use of protein identification technology to annotate and, if necessary, correct genomic DNA sequences. The third is the use of quantitative methods to analyse protein expression profiles as a function of cellular state as an aid to infer cellular function.

Proteomics, in particular quantitative proteomics, can be viewed as an array of biological or clinical assays capable of probing most, if not all, of the proteins in a sample. As proteins are involved in essentially all biological functions and clinical conditions, mass spectrometry and proteomics will have an even greater impact on biology and medicine than it has had so far.

PROTEOMICS APPROACHES FOR BIOMRKER DISCOVERY IN DISEASE

Despite tremendous advances in our understanding of the molecular basis of diseases such as cancer, substantial gaps remain both in our understanding of disease pathogenesis and in the development of effective strategies for early diagnosis and for treatment. The current interest in proteomics is due in part to the prospects that a proteomic approach to disease investigations will overcome some of the limitations of other approaches. The opportunities as well as the challenges facing disease proteomics are formidable. Particularly promising areas of research include: delineation of altered protein expression, not only at the whole-cell or tissue levels, but also in subcellular structures, in protein complexes and in biological fluids; the development of novel biomarkers for diagnosis and early detection of disease; and the identification of new targets for therapeutics and the potential for accelerating drug development through more effective strategies to evaluate therapeutic effect and toxicity.

Proteomic technologies can also be used to identify markers for cancer diagnosis, to monitor disease progression, and to identify therapeutic targets (Bensmail and Haoudi, 2003; Bensmail et al., 2005). Cancer proteomics encompasses the identification and quantitative analysis of differentially expressed proteins relative to healthy tissue counterparts at different stages of disease, from preneoplasia to neoplasia.

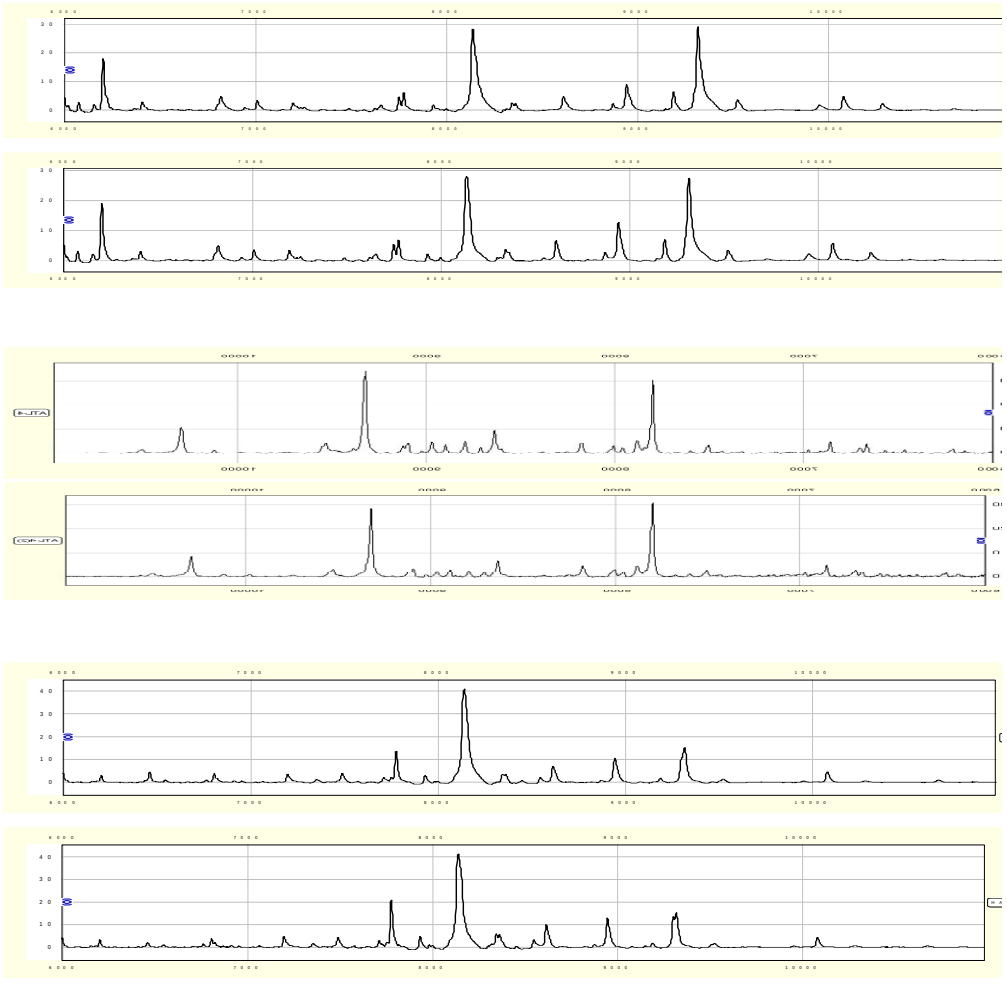

Figure 2: Representative spectra from Normal, ATL and HAM shown in duplicate from top to the bottom. These spectra illustrate the reproducibility of this MS-based proteomic approach in the analysis of biological samples.

Proteomics is valuable in the discovery of biomarkers because the proteome reflects both the intrinsic genetic program of the cell and the impact of its immediate environment. Protein expression and function are subject to modulation through transcription as well as through posttranscriptional and posttranslational events. There is substantial interest in applying proteomics to the identification of disease markers. Approaches include comparative analysis of protein expression in normal and disease tissues to identify aberrantly expressed proteins that may represent new markers, analysis of secreted proteins in cell lines and primary cultures, and direct serum protein profiling. The potential of mass spectrometry to yield comprehensive profiles of peptides and proteins in biological fluids without the need to first carry out protein separations has attracted interest. In principle, such an approach is highly suited for marker identification because of reduced sample requirements and high throughput.

This approach is currently popularized, particularly for serum analysis, by the technology referred to as surface-enhanced laser desorption/ionization. Microlitre quantities of serum from many samples are applied to the surface of a protein-binding plate, with properties to bind a class of proteins. There is currently a burgeoning interest in proteomics on the part of the pharmaceutical industry, evidenced by implementation of proteomics programmes by most major pharmaceutical companies.

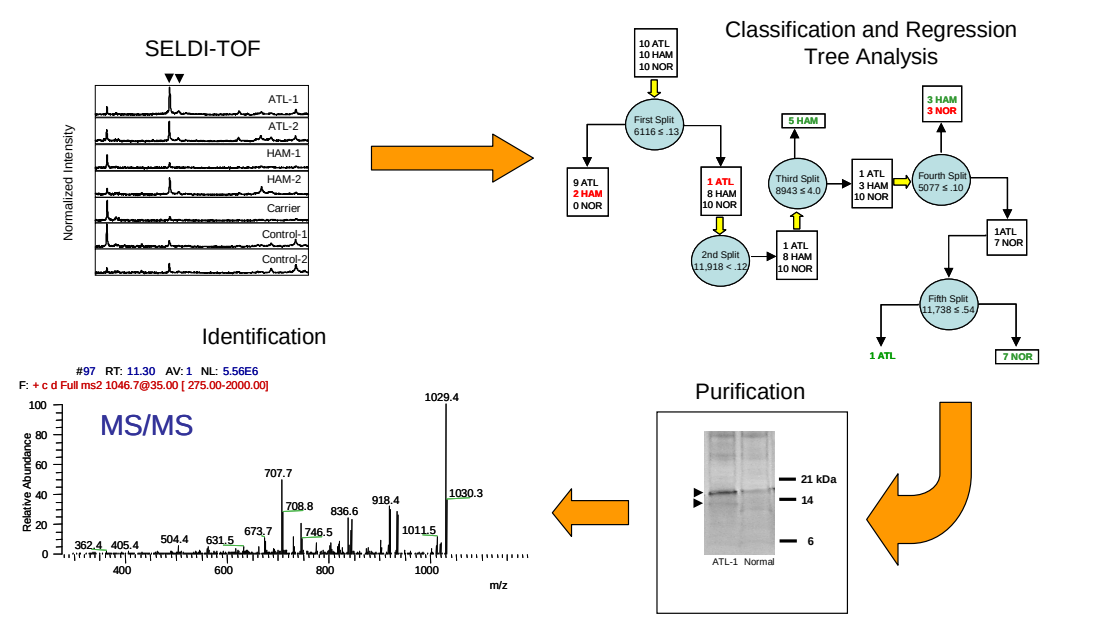

Figure 3: Summary of Biomarker Discovery and Identification of HTLV-1 –associated diseases biomarkers.

The notion has been advanced that, as the vast majority of drugs target proteins, proteomics should have substantial utility for drug development. But the industry has so far adopted a cautious attitude, and it is too early to make a critical assessment of the contributions of proteomics to drug development, relative to other approaches. The caution stems from the prior heavy investment in genomics and other approaches and some uncertainty surrounding the adequacy and scalability of proteomics to meet the needs of the pharmaceutical industry.

Provided suitable technology platforms become available, the use of proteomics may permeate numerous aspects of drug development, by identifying new targets and facilitating assessment of drug action and toxicity both in the preclinical and clinical phases.

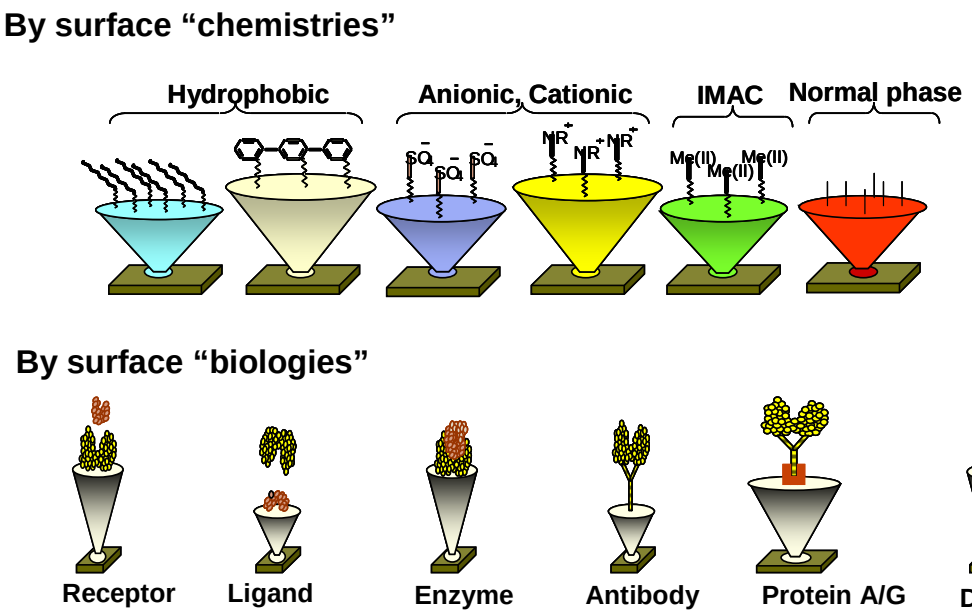

Figure 4: SELDI-protein-chips arrays available in different chemical and biological surfaces for different applications.

CONCLUSION

In studying a biological system using the biochemical approach, researchers have traditionally attempted to purify to homogeneity each of the system's components; each element is then studied in detail with the ultimate aim being to reconstitute the system *in vitro* from the isolated components. Because proteins carry out most biological activities, the biochemical approach has been significantly enhanced by the availability of the sensitive and rapid MS-based protein identification methods discussed in this article. The availability of complete genomic sequences from a number of species further facilitates MS-based protein identifications, as the requirement for *de novo* sequencing has been usurped by simple correlation of measured data versus theoretical data predicted from sequence databases. The availability of completely sequenced genomes also catalysed the emergence of systems biology — the attempt to systematically study all the concurrent physiological processes in a cell or tissue by global measurement of differentially perturbed states. The ultimate goal of systems biology is the integration of data from these observations into models that might, eventually, represent and simulate the physiology of the cell.

Proteome alterations in disease may occur in many different ways that are not predictable from genomic analysis, and it is clear that a better understanding of these alterations will have a substantial impact in medicine. A useful repertoire of proteomics technologies is currently available for disease-related applications, although further technological innovations would be beneficial to increase sensitivity, reduce sample requirement, increase throughput and more effectively uncover various types of protein alterations such as post-translational modifications. The use of these technologies will likely expand substantially, particularly to meet the need for better diagnostics and to shorten the path for developing effective therapy.

REFERENCES

Aebersold R. and Mann M. Mass spectrometry-based proteomics. *Nature* 422, 198 – 207 (2003)

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

Wright GL. SELDI proteinchip MS: a platform for biomarker discovery and cancer diagnosis. Expert Rev Mol Diagn. 2:549-563 (2002).

Bensmail H., and Haoudi A. Post-Genomics: Bioinformatics and Proteomics in Cancer Research. J. Biomed. Biotechnol. 4: 217-230 (2003)

Bensmail H., Golek, J, Semmes, OJ, and Haoudi A. Novel Approach for clustering proteomics data using Bayesian-based Fast Fourier Transform. Bioinformatics, 21(10): 2210-24. (2005)