

## New Developments in Proteomics

Mark W. Massing MD, PhD; Judyta Misiurek; Srinivas R. Chadaram, PhD; Christine E. Marx, MD, MA; Roger Madison, PhD

**G**enes are relatively static instruction sets for protein manufacturing processes in the cell. Fundamental genetic components (encoding regions) are linked, modified, and combined to create a wide variety of unique protein products. The total number of human protein-encoding genes has been estimated by the Human Genome Sequencing Consortium to be 20 000 to 25 000 genes.<sup>1</sup> The size of the proteome, the complete set of proteins expressed from the genome, is far larger and may exceed 100 000 proteins in humans.<sup>2</sup> Proteomics, the study of the proteome, is the next great challenge in biology and medicine and may rival genomics in complexity, costs, and benefits.

Legacy protein chemistry techniques such as chromatography, electrophoresis, and affinity columns have been used for decades and are an effective means to identify and characterize individual proteins. Proteomics is distinguished from protein chemistry in that proteomics tends to focus on patterns and systems of protein expression rather than on single components.<sup>3</sup> Proteomic techniques are capable

of simultaneously examining the expression of thousands of proteins to identify unique patterns associated with phenotypes, tissues, disease states, and responses to environmental or therapeutic exposures.<sup>4</sup> Clinical proteomics encompasses an understanding of protein systems in pathologic processes leading to new

diagnostic and prognostic tests, the discovery of protein targets for new pharmacologic therapies, and the identification of patients most likely to benefit from these therapies.<sup>5</sup>

The central problem in clinical proteomics is to distinguish and identify multiple proteins related to a disease or condition, even when these proteins are initially unknown. The underlying assumption is that a given disease or condition is manifested by a pattern of protein expression that is unique and identifiable. Proteomic methods compare protein expression in patients with and without a given condition to identify unique patterns or profiles of protein expression related specifically to that condition. Once a condition-specific protein expression pattern is discovered, its constituent proteins are identified

*“New and emerging technologies in the application of mass spectrometry to the field of proteomics offer clinicians a means to rapidly identify markers of disease leading to new diagnostic tests and treatments.”*

**Mark W. Massing, MD, PhD**, is a senior REAP fellow at the Research Service of the Veteran Affairs Medical Center in Durham, North Carolina. He can be reached at madis001@mc.duke.edu or 508 Fulton Street, Durham, NC 27705.

**Judyta Misiurek** is a senior research technician at the Research Service of the Veterans Affairs Medical Center in Durham, North Carolina, and attends the University of Miami Miller School of Medicine.

**Srinivas R. Chadaram, PhD**, is a research fellow at the Research Service of the Veterans Affairs Medical Center in Durham, North Carolina, and the Department of Surgery at Duke University Medical Center.

**Christine E. Marx, MD, MA**, is an assistant professor at the Research Service of the Veterans Affairs Medical Center in Durham, North Carolina and the Mid-Atlantic Mental Illness Research, Education, and Clinical Center and the Department of Psychiatry and Behavioral Sciences at Duke University Medical Center

**Roger Madison, PhD**, is a career research scientist at the Research Service of the Veterans Affairs Medical Center in Durham, North Carolina, and associate professor in the Departments of Surgery and Neurobiology at Duke University Medical Center.

as potential biomarkers for diagnosis and prognosis and as targets for treatment.

Mass spectrometry (MS) has become a promising technique in proteomics following advances supporting the processing of large molecules.<sup>4</sup> Mass spectrometry enables the separation and characterization of proteins in a complex tissue sample based on their different physical and chemical properties. The 2002 Nobel Prize in chemistry was awarded to John Fenn and Koichi Tanaka for their pioneering work in this area. Tanaka's approach utilized laser induced protein ionization and led to the development of matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) in the 1980s and to surface-enhanced laser desorption/ionization mass spectrometry (SELDI-MS) during the 1990s.<sup>6-8</sup>

In practice, these MS techniques are carried out in a series of steps. First, proteins are extracted from tissues by disruption of cellular structures and removal of nonprotein components. Next, protein solutions are cocrystallized with a matrix substance on specially developed chemically inert (in the case of MALDI-MS) or chemically active (in the case of SELDI-MS) surfaces. The matrix facilitates ionization of proteins when excited by laser energy.

Mass spectrometry devices identify patterns of protein expression by measuring the abundance of proteins at different molecular weights with a technique known as time-of-flight (TOF) detection. Figure 1 illustrates the basic concept of TOF detection. Proteins are ionized when struck by laser light and "fly" (ie, leave the surface). Ionized proteins are then captured by a high voltage electrical field and are accelerated in a vacuum chamber. During this acceleration period, or "flight," proteins become separated based on their charge and mass, arriving at a detector at different times. The more massive the protein, the less it is accelerated and the later it arrives at the detector.

Proteins striking a detector after TOF separation create a signal with an intensity related to the number of molecules arriving at the detector. The greater the abundance of molecules, the greater is the amplitude of the signal. Proteins with similar masses and charges arrive at the detector at approximately the same time creating a high

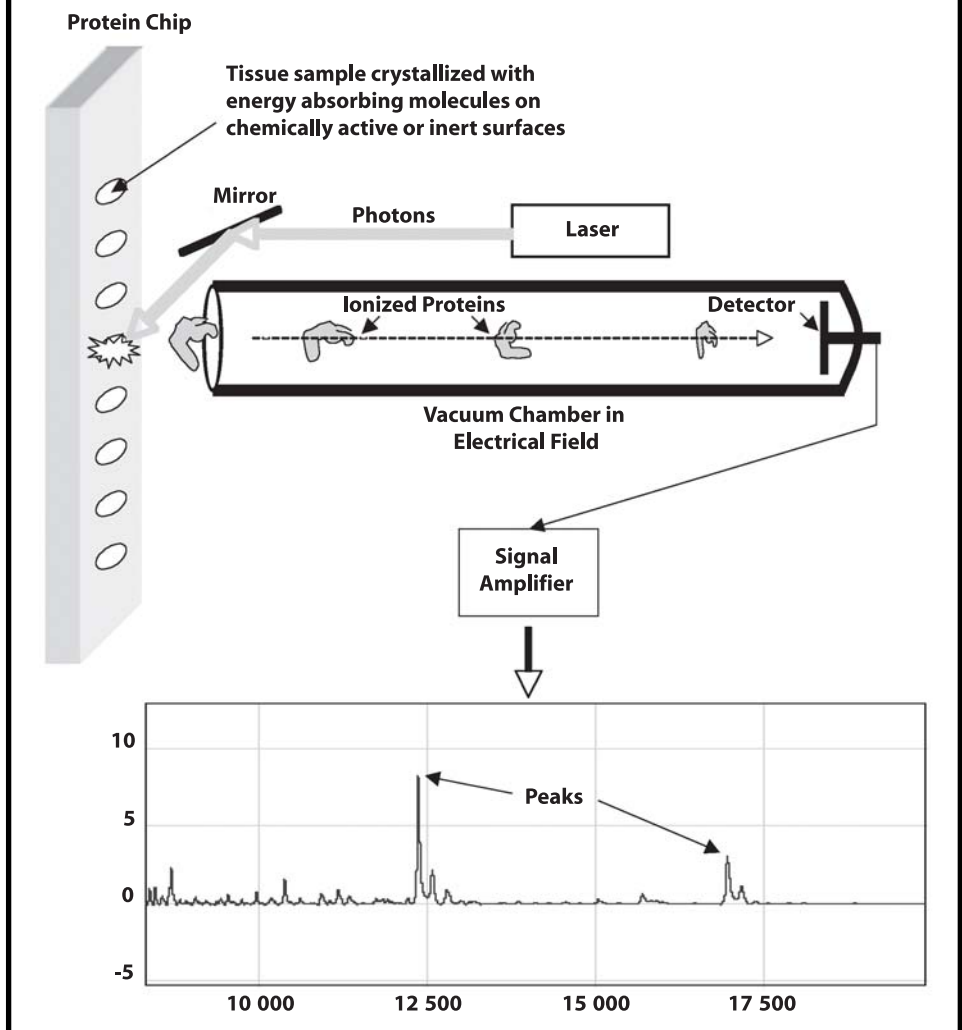
amplitude spectral "peak" (Figure 1). The pattern of peaks in a complex sample creates a spectrum—a unique fingerprint characterizing protein expression in a given tissue (Figure 2).

The MS spectrum graphically relates a protein's mass and charge (X-axis) to its abundance as measured by its signal intensity (Y-axis). Spectra from different tissue samples can be compared and common patterns of expression identified. Expression pattern differences can be mapped and analyzed. Peaks at similar mass-to-charge ratios (clusters) are identified across spectra (Figure 2) and relative signal amplitude differences are compared using sophisticated pattern recognition software to identify expression patterns that uniquely characterize specific diseases or conditions.

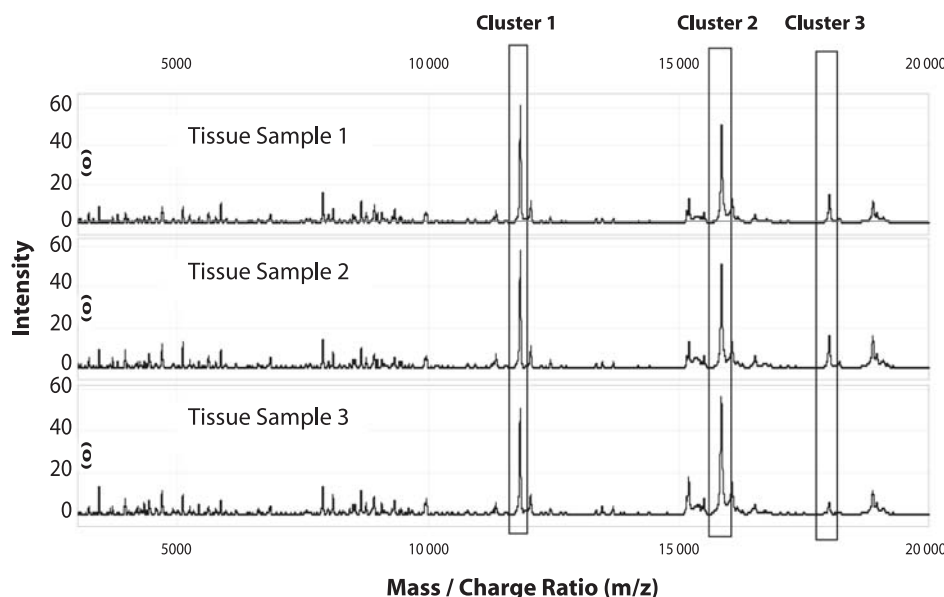
Recent developments in MS proteomics incorporate the use of chemically active surfaces on commercially available arrays known as protein chips.<sup>9,10</sup> Chemically active surfaces allow for on-chip selective extraction of proteins based on chemical properties to simplify processing of complex clinical samples.

Despite the promise of this new technology, a number of

**Figure 1.**  
**Basic Components of a Laser Desorption-ionization Mass Spectrometry System Used in Clinical Proteomics.**



**Figure 2.**  
**Spectra from Three Samples of Rat Quadriceps Muscle Showing Three**  
**Signal Peaks Sharing Mass-to-Charge Ratio Values (Clusters).**



technical obstacles impede its rapid adoption. Often the most difficult issue is defining and obtaining clinical samples suitable for proteomic analyses. Variations in patients, sample handling, and collection protocols constitute substantial challenges. The first step in any proteomics experiment is to obtain and prepare the tissue sample for processing. Tissue preparation is frequently the most resource intensive activity.

Investigators are currently developing and refining SELDI-MS protocols to process a variety of tissue types including serum/plasma, brain, cerebrospinal fluid, urine, tears, saliva, cells from washes and biopsies, and muscle. The earliest efforts at biomarker discovery with SELDI-MS focused on markers associated with various cancers, especially those remaining asymptomatic until late stages such as ovarian<sup>11</sup> and pancreatic<sup>12</sup> cancers. Diagnostic and prognostic tests for these diseases were desirable and tissue preparation protocols for blood were developed early and have been refined considerably over the years.<sup>13</sup>

An issue of critical importance for proteomics analysis of complex biological and clinical samples for discovery of biomarkers is the need for reduction of tissue sample complexity prior to MS analysis. Most tissue samples contain far too many proteins to be evaluated on a single protein chip. These complex samples are broken down into a series of less complex fractions based on the chemical properties of constituent proteins. Conventional methods such as fractionation of complex clinical samples by ionic exchange chromatography and new methods such as enriching low abundant proteins by affinity capture with a combinatorial library of ligands<sup>14</sup> provide much needed tools for processing complex biological and clinical samples for proteomics research.

Another major concern is ensuring that tissue preparation and subsequent processing is standardized and does not vary

between samples within experiments. Tissue samples from different individuals are never uniform. Even if gross tissue mass is identical, differences in connective tissue, vascularization, and fat content may result in differences in tissue protein expression patterns.

In examining the entire proteome, it is frequently the case that multiple protein expression differences are found when comparing tissues from different sources or time frames. A challenging problem in proteomics is the identification of patterns of expression associated with a given condition of interest using voluminous experimental data.<sup>15</sup> Mass spectrometry analysis of the

proteome can generate an intimidating amount of data. A single clinical tissue sample could generate many thousands of data points describing protein expression patterns. Even small experiments generate too much data to be processed manually. A variety of different approaches, frequently borrowed from genomics, have been used including decision tree analyses, genetic algorithms, and neural networks.<sup>16-18</sup> Development of standardized and universally accepted approaches to analyze protein expression patterns is a goal that has yet to be realized.

It is hard to overstate the potential clinical relevance of the application of MS to the field of proteomics. New and emerging technologies offer clinicians a means to rapidly identify markers of disease leading to new diagnostic tests and treatments. Objective screening tests for conditions such as psychiatric illness based on proteomic techniques could revolutionize the care of patients and lead to better treatments. However, it is important to temper our enthusiasm with an understanding of the challenges that await us as nascent proteomics technologies mature. Sound experimental protocols and analytic methods must keep pace with the rapid development of proteomics tools and hardware. A rush to process experiments without considering common standards and potential pitfalls could generate misleading results and wasted effort. With this caveat in mind, the upcoming era of proteomics should complement genomics and provide a direct clinical relevance not possible by genomics alone. **NCMJ**

*Acknowledgement: This work was made possible by a Research Enhancement Award Program (REAP) to the Veterans Affairs Medical Center in Durham, NC from the Biological Laboratory and Research Service of the Office of Research and Development of the Department of Veterans Affairs (RM principal investigator).*

## REFERENCES

- 1 Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431:931-945.
- 2 Harrison PM, Kumar A, Lang N, Snyder M, Gerstein M. A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res*. 2002;30:1083-1090.
- 3 Liebler DC. *Introduction to Proteomics: Tools for the New Biology*. Totowa, New Jersey: Humana Press, Inc.; 2002:1-24.
- 4 Walker SJ, Xu A. Biomarker discovery using molecular profiling approaches. *Int Rev Neurobiol*. 2004;61:1-30.
- 5 Petricoin E, Wulfkuhle J, Espina V, Liotta LA. Clinical proteomics: revolutionizing disease detection and patient tailoring therapy. *J Proteome Res*. 2004;3:209-217.
- 6 Issaq HJ, Veenstra TD, Conrads TP, Felschow D. The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification. *Biochem Biophys Res Commun*. 2002;292:587-592.
- 7 Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*. 1988;60:2299-2301.
- 8 Hutchens TW, Yip TT. New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom*. 1993;7:576-580.
- 9 Tang N, Tornatore P, Weinberger SR. Current developments in SELDI affinity technology. *Mass Spectrom Rev*. 2004;23:34-44.
- 10 Reddy G, Dalmasso EA. SELDI ProteinChip(R) Array technology: Protein-based predictive medicine and drug discovery applications. *J Biomed Biotechnol*. 2003;4:237-241.
- 11 Zhang Z, Bast RC, Yu Y, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res*. 2004;64:5882-5890.
- 12 Koopmann J, Zhang Z, White N, et al. Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. *Clin Cancer Res*. 2004;10:860-868.
- 13 Petricoin EF, Liotta LA. SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. *Curr Opin Biotechnol*. 2004;15:24-30.
- 14 Huang HL, Stasyk T, Morandell S, et al. Enrichment of low-abundant serum proteins by albumin/immunoglobulin G immunoaffinity depletion under partly denaturing conditions. *Electrophoresis*. 2005;26:2843-2849.
- 15 White CN, Chan DW, Zhang Z. Bioinformatics strategies for proteomic profiling. *Clin Biochem*. 2004;37:636-641.
- 16 Li L, Umbach DM, Terry P, Taylor JA. Application of the GA/KNN method to SELDI proteomics data. *Bioinformatics*. 2004;20:1638-1640.
- 17 Qu Y, Adam BL, Yasui Y, et al. Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. *Clin Chem*. 2002;48:1835-1843.
- 18 Ball G, Mian S, Holding F, et al. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. *Bioinformatics*. 2002;18:395-404.