An introduction into proteomics and its clinical applications

Mamoun Ahram, PhD.

ABSTRACT

The completion of the human genome project has marked the official start of the post-genomic era. Due to limitations in DNA and RNA studies, the advent of proteomics, or large-scale analysis of proteins, is considered a crucial consequence and a chief player of post-genomic initiatives. An immediate goal of proteomic studies is the understanding of proteins including their expression, function, interaction, and structure with an endpoint of discovery of protein biomarkers. Such biomarkers can be used in detection, prognostication, and treatment of diseases. However, the challenge of studying complete sets of cellular proteins, or proteomes, is driving the development of newer technologies. In this review, a discussion of proteomic biotechnologies and their clinical applications is offered. The main techniques covered are the 2-dimensional polyacrylamide gel electrophoresis, mass spectrometry, surface-enhanced laser desorption ionization mass spectrometry, and protein microarrays. These tools have been successfully utilized in analyzing human and other biological samples in discovery of biomarkers. Recent advancement in proteomics has added and will continue to add valuable information to our knowledge-base of the human biological system.


From the Department of Pharmacology and Physiology, Faculty of Medicine, Mutah University, Mutah, Jordan.

Received 14th August 2006. Accepted 16th December 2006.

Address correspondence and reprint request to: Dr. Mamoun Ahram, Department of Pharmacology and Physiology, Faculty of Medicine, Mutah University, Mutah, PO Box 7, Karak 61710, Jordan. Tel. +962 (79) 5567779. Fax. +962 (3) 2386175. E-mail: Dr.ahram@gmail.com

The completion of the human genome project has generated tremendous excitement in the scientific community with the many opportunities it provides not only in unraveling the human genome, but also in understanding the complexity of the human biological system as a whole. However, many deficiencies exist in regards to the basic knowledge of human genes. First of all, of the 20,000-30,000 genes that make up the human genome, the expression pattern in different cells and tissues has yet to be determined. It has been estimated that ~6,000 genes are expressed per cell with a cell specificity of ~400 proteins. This estimate is reached without considering the protein end-products following post-translational modifications. In addition to lack of knowledge in regards to gene expression, the function of many of the human genes has not been elucidated. A closer examination of a human protein database reveals the presence of a considerable fraction of hypothetical gene products, which has been estimated to be approximately 60% of a database.

Major modern scientific research efforts are aimed towards “molecular profiling”, or global measurements of mRNA and proteins levels in biological systems. This strategy is based on information provided by the human genome project and aided by development of technologies for molecular analysis of diseases. Current technologies used in molecular profiling studies generate enormous amount of data points in a short period of time, thus termed “high-throughput.” High-throughput studies allow the concurrent investigation of thousands of genes enhancing the discovery of novel genes and elucidation of molecular and cellular interactions. However, due to the relative biochemical consistency of RNA molecules and the ease of mRNA manipulation in contrast to proteins, molecular profiling studies and technological developments have mainly been driven towards determination of mRNA levels.

Although RNA-based studies are promising in identifying gene products linked to diseases, such studies suffer from serious shortcomings. First, it is important to realize that mRNA molecules are transit products into making proteins, the primary functional player in cells. In addition, RNA transcript levels may not correspond to real protein levels due to differences in regulation and stability. This has been illustrated in previous studies where levels of mRNA molecules and their corresponding proteins poorly correlate. In addition, proteins may undergo extensive modification such as proteolysis, glycosylation,
or phosphorylation generating different isoforms each with distinct function. The same protein may also differ in its localization in physiological and pathological conditions influencing its function. Finally, proteins are the main target of therapeutic agents accounting for more than 98% of drug targets. Due to the biological significance of proteins, the field of “proteomics” has launched.

Many attempts have been made to define the term, “proteomics.” A simple definition can be stated as “high-throughput analysis of proteins” involving hundreds to thousands of proteins. Among the many goals of proteomics is understanding of all aspects of proteins including their expression, function, molecular interaction, and structure. It is hoped with proteomic analyses to discover novel disease biomarkers that can be utilized for detection, prognostication, and treatment of various diseases. The challenge of studying complete sets of proteins in cells, or proteomes, has been driving the development of new technologies. In fact, proteomics can be considered an application of evolving biotechnologies. Continuous improvement of older instruments has facilitated the execution of prominent proteomic studies. In addition, newer and powerful instruments have also been developed. In this review, commonly used proteomic technologies and examples of their applications are introduced. The clinical aspect of these techniques will also be emphasized. As will be noted, no proteomic technique will reveal the mystery of our proteomes and, therefore, multiple methods are needed to achieve an appreciable level of knowledge.

**Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE).** Two-dimensional polyacrylamide gel electrophoresis is rather an older protein analytical tool invented more than 30 years ago. However, since mid-1990’s, it has been revived and been the proteomic tool of choice in many laboratories (for review). This technique enables the separation of proteins in 2-dimensions based on their isoelectric point (first dimension) followed by their mass (second dimension) (Figure 1). A few hundreds to approximately two thousands of proteins can be visualized by 2D-PAGE producing a “protein fingerprint” of cells. Major improvements have been introduced making 2D-PAGE technically user-friendly and reproducible. For example, the development of immobilized pH gradient (IPG) strips has allowed consistent and reproducible focusing of proteins in the first dimension. In addition, protein focusing can be performed at different ranges of isoelectric points such as a standard 3-10 range or narrower ranges including 4-7, 5-8, and 6-11, allowing wider separation and, hence, the visualization more proteins.

Once separated in the second dimension by a standard polyacrylamide-based gel electrophoresis, proteins can then be visualized using a staining dye. Currently, there are 3 types of staining procedures commonly used for detection of protein spots. The first is based on a blue dye known as Coomassie blue. This relatively inexpensive procedure is simple and allows the detection of proteins as low as 1 ng. The second staining method is based on silver nitrate requiring more elaborate work than the former method, but is more sensitive where proteins as low as 0.1 ng can be detected. However, due to the utilization of a fixative in the silver staining method, the sensitivity of protein identification by mass spectrometry (to be covered later) is significantly compromised. As a result, a fluorescent method using a dye known as Sypro has been reported providing the same sensitivity of silver staining and the simplicity of Coomassie blue staining.

A drawback of fluorescent dyes, however, is their high expense relative to the other 2 methods. It is now possible to analyze 2 samples, namely, normal and diseased, by the same 2D gel with the commercial introduction of 2-dimensional differential in-gel electrophoresis (DIGE). This technique overcomes problems associated with comparative quantification of protein spots in two different gels due to preparative and staining artifacts. In DIGE, the proteins of 2 samples are labeled with different fluorescent tags. Then, they are combined at equal amounts and separated by 2D-PAGE as is normally performed for a single-sample experiment. The gel can be visualized under a special imaging system where each sample is scanned independently and the images are superimposed for comparative, computational analyses. DIGE has been utilized in discovery of protein biomarkers in breast and esophageal cancers.

Celis et al. have illustrated the occurrence of major modifications in protein expression patterns when cells are removed from their natural in vivo environment. Similar observations are made when the protein profile of prostate cells isolated from a tissue sample is compared to that of the same cells but grown in vitro. Thus, direct proteomic studies of human tissues are critical, in particular, since tissues are the true representation of physiological and pathological states. The advent of Laser Capture Microdissection (LCM; Arcturus Engineering, http://www.arcturus.com/) has made it possible to study protein profiles of pure populations of cells procured from tissue sections (Figure 2). Laser Capture Microdissection has been successfully utilized in discovery of protein biomarkers including those
of prostate and esophageal cancers.\textsuperscript{21,22} The source of tissues has been shown to not only be limited to frozen samples, but can also include ethanol-fixed, paraffin-embedded tissues.\textsuperscript{22,23}

Although 2D-PAGE is an extremely powerful technique, it suffers from major shortcomings. These include the inability to resolve hydrophobic as well as highly basic proteins. In addition, 2D-PAGE can detect a small fraction of a cellular proteome. For example, if a cell consists of 6,000 proteins as hypothesized,\textsuperscript{1} 2D-PAGE can detect no more than one-third of the proteome in perfect circumstances. This is a significant limitation if we consider that intermediate and low-abundance proteins such as protein kinases and growth factors are major players in human diseases. 2D-PAGE is also limited to analyzing proteins larger than 5 kDa.

One dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (1D-SDS-PAGE).

Due to limitations in analysis of certain protein classes by 2D-PAGE, the traditional 1D electrophoretic method, namely SDS-PAGE, has been employed in a number of proteomic studies. Sodium dodecyl sulfate is an ionic detergent that has proven superior in general protein solubilization. In this technique, proteins are resolved based on mass only. For large-scale proteomic studies, the gel lane containing separated proteins is cut into smaller pieces and proteins in each piece are identified by mass spectrometry. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis has led to the identification of numerous proteins from membrane preparations of colon cancer cells\textsuperscript{24} and breast cancer.\textsuperscript{25} For review on proteomic analysis of membrane proteins, refer to Ahram and Springer.\textsuperscript{2} However, similar to 2D-PAGE, SDS-PAGE is limited to analyzing proteins larger than 5 kDa. In addition, due to resolving and staining limitations of SDS-PAGE, it is not practical to perform visual comparison of individual proteins, especially for crude samples.

Mass spectrometry (MS).

Currently, MS is currently the method of choice in high-throughput proteomic studies. A mass spectrometer measures the masses of small molecules by converting them into ions and sorting

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{Protein analysis by 2-dimensional polyacrylamide gel electrophoresis. Proteins are separated by in 2 dimensions based on 2 biochemical features: isoelectric point or charge (first dimension) and mass (second dimension). In the illustrated gel, proteins are separated at a range of pH 3-10 in the first dimension and 5-100 kDa in the second dimension. Following staining, each spot represents at least one protein. The same protein may be of various isoforms that differ in charge, but not in mass due to post-translational modifications. A protein spot can be excised from the gel for identification by mass spectrometry as illustrated in Figure 3.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig2.png}
\caption{Precise procurement of cell populations by LCM. a) Prostatic glandular epithelial cells in an ethanol-fixed, paraffin-embedded tissue section are imaged before isolation; b) Some of these cells are procured with high accuracy and specificity; c) the remaining of the tissue following procurement is intact and neighboring cells can be procured for comparative analysis.}
\end{figure}
them via a stream of electrical fields according to their mass/charge (m/z) ratio. Mass spectrometry instruments are composed of 3 components: an ionization source, a mass analyzer, and an ion detector. Recent and continuing development and improvement of these instruments have revolutionized the field of proteomics. In particular, a number of modifications have enabled high-throughput analysis of protein samples including the invention of soft ionization methods, the innovation of hybrid instruments composed of 2 mass analyzers, and the direct integration of liquid chromatography to MS instruments.

Prior to protein analysis by MS, the protein is specifically digested into smaller peptide fragments by a protease, preferably trypsin. The sample is then introduced into the mass spectrometer for ionization. There are 2 common types of “soft” ionization methods; the first is termed as Matrix-Assisted Laser Desorption/Ionization (MALDI), where samples are embedded into specific matrix molecules. The matrix absorbs the ionization laser beam and transfers the energy into the analyte. The second ionization method was termed as electrospray ionization (ESI) in which peptides are injected into the ionizing chamber to be converted into smaller droplets. These droplets are vaporized in the presence of an electric field creating charged molecules. Once ionized, molecules are directed via a mass analyzer towards a detector. Whereas ESI requires the peptides be in liquid state and generates ions of multiple charges, MALDI can handle peptides in gaseous, liquid, or solid state and generates ions of similar amounts of energy. In addition, MALDI is more tolerant than ESI to contaminants such as salts and non-ionic detergents.

In addition to being classified according to the ionization method, namely ESI or MALDI, MS instruments can be classified according to their component of mass analyzer. Among the most commonly used analyzers are ion trap (IT), quadrupole (Qq), Fourier-transform ion cyclotron resonance (FTICR), and time-of-flight (TOF). In TOF, the ions are accelerated from the ionization source down a flight tube until they impact the ion detector at the other end of the tube. Since all ions carry the same amount of energy but have different masses, smaller ions travel faster and reach the detector earlier than larger ones. Based on the time needed for the ion to reach the detector, the masses on ions can be calculated generating a “peptide mass fingerprint” (PMF) of the analyzed protein. Time-of-flight mass analyzer is commonly combined with MALDI instruments, hence known as MALDI-TOF.

Electrospray ionization is usually interfaced to other mass analyzers including ion trap, quadrupole, and FTICR. Quadrupole mass analyzers are composed of 4 parallel rods generating an electrical field that focuses the ions and sorts them according to their m/z. In ion trap analyzers, ions are trapped in a 3-dimensional electrical field for selective isolation. Fourier-transform ion cyclotron resonance is a special type of ion traps except that ions are trapped in a magnetic field rather than an electrical one. Fourier-transform ion cyclotron resonance is a powerful mass analyzer providing the highest sensitivity, resolution and mass accuracy. For example, it has been reported that FTICR-MS can identify peptides as low as zeptomoles (10-21 moles).

Since proteins would differ in their proteolytic digestion generating unique PMF, the protein identity can be determined. This is carried out by comparing the peptides’ mass spectra of the analyzed protein to theoretical mass spectra of other proteins in a sequence.
Introduction into proteomics...Ahram

Figure 4 - Representation of protein profiling by surface enhanced laser desorption ionization (SELDI). Measurement of protein time-of-flight can be illustrated graphically as a chromatogram peaks (A), gel-like bands (B), or a mass spectra (C) with every peak, band, or spectrum correspond to a protein.

Surface enhanced laser desorption ionization (SELDI). Surface enhanced laser desorption ionization is a new technology that is a modification of MALDI-TOF mass spectrometry. A major difference is that a sample is applied on surface of a chip rather than mixed with a matrix molecule. The chip surface is made of a defined chemical property (such as hydrophobic, cationic, and anionic) allowing certain classes of proteins to adsorb. The chip is then placed in a vacuum chamber of the SELDI instrument where proteins and peptides are ionized and travel towards a detector inversely according to their masses. The time-of-flight of ions, in particular of peptides below 20 kDa, can be viewed as MS spectra, a chromatogram, of gel-like bands (Figure 4).

Due to its speed, high sensitivity, ease of use, and reproducibility, SELDI is a true high throughput proteomic instrument. Two major advantages of SELDI is the ability to analyze highly complex samples, and the low volume of needed for analysis, which can be as low as 0.5 µl. SELDI is also versatile where, instead of a chemical surface, the chip can be coated with antibodies to capture specific antigens has been reported earlier in measuring prostate-specific antigen and prostate-specific membrane antigen.36-38

Surface enhanced laser desorption ionization has also been utilized in search for biomarkers for Alzheimer’s disease39,40 as well as cancers of the prostate,41-43 bladder,44 colon,45 and breast.46 Although direct determination of proteins represented as mass peaks is not possible, different means can be utilized to reveal the identity of specific peaks. In a recent report, SELDI analysis of cerebrospinal fluid (CSF) samples of patients with multiple sclerosis revealed the presence of a differential...
A differential peak was observed between control and multiple sclerosis samples. This protein was identified by further MS analyses as cystatin C, an inhibitor of the lysosomal cysteine protease cathepsin B. Although burdensome and elaborate, proteins represented by specific SELDI spectra peaks can be identified by a series of liquid chromatography fractionation has been illustrated by Diamond et al, Sanchez et al, and Yang et al.

In one study, proteins extracted from LCM-microdissected prostatic normal and tumor cells were analyzed by SELDI. The mass spectra patterns of the proteins revealed several remarkable alterations as compared to those of matched normal samples. However, due to the dynamic heterogeneity of proteomes even within the same individual, consistent detection of differential peaks is not feasible. This complexity has prompted the group of Petricoin and Liotta to integrate an artificial neural network algorithm to search for “hidden” patterns. In a fascinating study, they have been able to differentiate ovarian cancer patients from normal subjects and patients with other ovarian diseases with unprecedented sensitivity of 100% and specificity of 95%. These results are significant accomplishment considering that, first, the 5-year survival rate is 11% for stage IV ovarian cancer patients in comparison to up to 93% for early-stage patients, and second, there are no superior biomarkers for detecting early-stage ovarian tumors, and, therefore, two-third of ovarian cancers are detected in advanced stage. Very importantly, the SELDI analysis is performed using a small volume (a few microliters) of unfractionated serum samples. The latter study has initiated similar studies on numerous diseases with promising findings in detecting a variety of disorders.

**Protein microarray.** The success of DNA microarrays has encouraged scientists to invent a similar technology for proteins. Such technology, known as protein microarrays, would allow high-throughput, high sensitivity, and robust analysis of thousands of proteins. Different types of protein microarrays have been designed and can be categorized according to their end-point purpose (Figure 5). The first most common design is termed expression or abundance–based protein microarrays. This type of microarrays entails spotting thousands of bait molecules on a glass or membrane-coated slide. Each spot would then represent specific bait for a single protein. Usually the bait molecule is an antibody, although other capture molecules such as aptamers (small DNA or RNA molecules) or phage lysates can be used. By incubating a sample containing mixed populations of proteins onto the spotted slides, protein molecules would bind specifically to the corresponding bait molecule. Captured proteins can be detected by directly labeling the sample before applying them onto the slide. This direct labeling method has been utilized in identifying biomarkers of prostate cancer and radiation-regulated proteins. With the direct labeling method, it is possible to analyze more than one sample each labeled with a different fluorescent tag. Otherwise, an indirect labeling method, also known as sandwich immunoassay, can be used where bound proteins are targeted by a second bait such as a different primary antibody that targets a different domain as has been illustrated in measuring the expression the epidermal growth factor receptor and ERB2 and monitoring EGF-dependent phosphorylation in human tumor cells.

Both direct and indirect labeling expression microarrays are also known as forward-phase protein
microarrays. They are in contrast to another design of expression microarray technology termed “reverse-phase protein microarrays”. This technology involves spotting the analytes (namely protein extracts) rather than bait molecules, with each spot representing a single test sample. Lysates can be spotted at different dilutions providing an internal standard curve and an opportunity for quantitative measurement. It is worthy to mention that spotted lysates can be obtained from LCM-microdissected cells allowing for studies of pure cell populations. For example, differential protein expression in microdissected prostate cancer cells has been compared to that in patient-matched normal and premalignant cells from the same tissue samples. In addition to expression pattern, signal transduction circuitry can be studied using reverse-phase protein microarrays. Whereas with forward-phase proteins microarrays, one can analyze multiple proteins in a maximum of 2 samples per one microarray, reverse-phase protein microarrays enable investigation of one or 2 proteins in multiple samples per one microarray. Both types of expression protein microarrays are hampered, though, by the availability of a specific bait molecule (antibody, for example).

Since a major goal of proteomics is to assign function to proteins, 2 other types of protein microarrays have been developed: functional microarrays and interaction microarrays. The enzymatic activities of purified 119 yeast proteins predicted to be tyrosine kinases have been investigated using a functional microarray where kinase substrates are arrayed on a solid surface. Purified proteins are then added individually to the microarrays in the presence of Py-ATP. The ability of 27 of the 119 proteins to phosphorylate certain substrates has confirmed their kinase activity. In addition, effects of DNA mutations and polymorphism on DNA-protein binding have been studied using interaction protein microarrays. Interaction arrays can be designed to investigate interaction of certain proteins to various types of molecules including other proteins, peptides, nucleic acids, lipids, carbohydrates, and small molecule.

The future. Despite the many challenges, it is certainly an exciting time, the scientific research is going through. Completion of the Human Genome Project has surely been a major factor in increasing understanding of many aspects of biological systems. Similarly, knowledge of human proteomes will contribute significantly to elucidating the normalcy as well as the malfunctions of our genetic and proteomic makeup. It is, however, integration of genomic and proteomic information that will offer a clearer picture of the human biological system. This knowledge will help us understand the physiological and pathological conditions and, hence, the designs are safer and more specific therapeutics. We hoped that we achieve the level of personalized medicine, where therapy is tailored according to the genetic and proteomic makeup of individuals. In order to achieve these goals, stronger interaction and collaboration are needed among clinicians and basic scientists.

References

Introduction into proteomics ... Abram


